REMARKS

The present invention relates in part to the use of affinity tags in recombinant fusion protein constructs. In particular, the claimed invention relates to affinity tags which comprise two or more modules capable of mediating binding to streptavidin.

Claims 1–34, 36, 37, 40–45 and 47-51 are pending in the application, with claims 16, 17, 32–34, 36, 37, 40–45, and 47-51 under examination. The balance of the claims have been withdrawn from examination by the Examiner in accordance with a restriction requirement.

Reconsideration of the claimed invention is respectfully requested in view of the foregoing amendments and the remarks contained herein.

I. Rejection Under 35 U.S.C. §103

The rejection of claims 16, 17, 32-34, 36-37, 42-45, and 47-51 under 35 U.S.C. §103(a) as allegedly being obvious over Skerra and Schmidt, *Biomedical Engineering* 16: 79-86, 1999, in view of Srisawat *et al.*, RNA 7: 632-41, 2001, is respectfully traversed.

A. The secondary Srisawat *et al.* publication is not citable as prior art

Applicants note that the present application claims priority from German patent application 20011013776, having a priority date of March 21, 2001. The publication date of the secondary Srisawat *et al.* publication is July 1, 2001, and is thus not citable as prior art against the present application.

The Office Action asserts that the cited references are citable under 35 USC 102(e); however, neither cited reference is a patent application eligible for citation under 35 USC 102(e). Applicants request clarification of the basis on which the publications, and in particular Srisawat *et al.*, is being cited against the pending claims.

While the initial burden of establishing a rejection lies with the examiner, Applicants provide the following comments to aid the Examiner in further analyzing the citability of the secondary Srisawat *et al.* publication. The pending claims are fully supported by the priority document (German patent application 20011013776), an English translation of which is provided herewith, and by PCT/EP01/11846, which was filed 12 October 2001. It is noted that in order to overcome a rejection under 35 U.S.C. 103, Applicants must show possession of something falling within the claims prior to the effective date of the reference being antedated. MPEP 715.02. In this regard, Applicants point to pages 6 and 7 of the German priority application which discloses the use of two sequentially arranged streptavidin binding peptide modules; claims 2 and 3 of the German priority application, page 6, lines 16 to 21 and page 14 lines 20 to 24 which disclose the sequences His–Pro–Baa and –His–Pro–Gln–Phe–; and page 16, last paragraph and claim 10, for example, that discloses a respective fusion protein.

In the event that the rejection is maintained and the basis under which the publications, and in particular Srisawat *et al.*, is being cited against the pending claims is clarified, Applicants reserve the right to file an affidavit under 37 CFR 1.131 if considered appropriate. However, based on the German priority application and PCT/EP01/11846, Applicants respectfully submit that it is apparent from the record that the secondary Srisawat *et al.* publication is not citable as prior art against the present application.

B. Even if the prior art was citable against the present claims, the claims are not obvious over Skerra and Schmidt in view of Srisawat *et al.*

The claimed invention relates to provision of a fusion protein comprising a streptavidin–binding peptide linked to a protein sequence of interest. As recited in claim 16, the streptavidin–binding peptide comprises a sequential arrangement of two modules with an amino acid sequence of –His–Pro–Baa– in which Baa is selected from the group consisting of glutamine, asparagine and methionine. At least one of the modules comprises a sequence –His–Pro–Gln–Phe–.

Applicant notes that Thomas Schmidt, the inventor named in the present application, is also named as an author of the cited Skerra and Schmidt article. This article is also summarized in paragraphs [0008] to [0011] of the present specification.

The primary Skerra and Schmidt publication discusses two commercially available peptide sequences having binding properties towards streptavidin, invented by Drs. Skerra and Schmidt prior to the present application (*see, e.g.*, U.S. Patent 5,506,121, also discussed in paragraphs [0008] to [0011] of the present specification). These two sequences are:

- *Strep-tag*: Ala-Trp-Arg-His-Pro-Gln-Phe-Gly-Gly (page 80, right column, top of page); and
- *Strep-tag II*: Asn-Trp-Ser-His-Pro-Gln-Phe-Glu-Lys (page 81, left column, first full paragraph).

While Skerra and Schmidt discuss the use of these previously known peptide affinity tags individually, nowhere does the article discuss any sequential arrangement of two such modules, as required by the present claims.

The secondary Srisawat *et al.* publication is cited for allegedly recognizing "that affinity tags can be used in tandem and in multiples." Office Action, page 4. In fact, the secondary Srisawat *et al.* publication never discusses the use of multiple streptavidin binding peptides as the Office Action asserts. Srisawat *et al.* discloses certain RNA aptamers and fusions of individual aptamers with RNA molecules. The length of each aptamer is about 45 nucleotides in length, and the binding site of these RNA aptamers on streptavidin is not known. The underlined portions of the aptamers shown in Fig. 1 of Srisawat *et al.* does not depict tandem aptamers; it depicts the conserved regions of two aptamer families. Srisawat *et al.* does not teach the use of multiple streptavidin binding peptides; therefore even if combined with Skerra & Schmidt, the combination of publications does not lead to the claimed invention.

It is further asserted that a motivation to combine Srisawat *et al.* with Skerra and Schmidt article is because "increased numbers of tags would enhance affinity." Office Action, page 4.

This assertion for providing "increased numbers of tags," however, is incorrect and does not properly reflect the physics and chemistry of how binding affinities are understood in the art.

The affinity of a protein to its ligand is a defined by the association constant K_a (or the reciprocal value thereof, the dissociation constant K_d). As stated in paragraph 8.2.1 on page 338 of Creighton, Proteins, Structure and Molecular Properties, 2nd Edition, 1993, W.H. Freemann and Company, New York attached hereto "K_a is a constant under a given set of conditions and is measured experimentally by the dependence of binding on the free ligand concentration." Simply increasing the number of tags does not alter the affinity in any way – the affinity of each tag for its ligand, and the overall affinity of the multiple tags, does not change.

Even assuming the assertion is correct, however, one of skill in the art would not be motivated to increase the number of tags as the Office Action suggests. The intended purpose of Srisawat *et al.* as stated on page 632 of the publication is "to develop an RNA tag that binds tightly to a commonly available target molecule in such a way that the ligand–RNP complex can be selectively and gently dissociated afterwards." Driving the affinity ever higher would require increasingly harsh dissociation conditions. Ultimately, if it were correct that increasing the number of tags increases affinity, the logical conclusion is that increasing the number of tags would lead to a ligand-RNP complex which cannot dissociate without destroying the proteins of the complex.

Moreover, the Office Action fails to consider in any way the substantial evidence of nonobviousness already of record in the present case. As discussed by Dr. Schmidt in paragraphs 8-11 of his previously submitted declaration, the present invention is drawn to improved fusion proteins and peptides able to selectively bind to substrates comprising streptavidin (which term means either or both naturally-occurring and mutein or optimized forms such as the commercially available STREP-TACTIN® peptide). The streptavidin-binding peptides of the present invention have the ability to bind strongly to their substrate under non-competitive conditions and yet may be easily displaced under competitive conditions. This bimodal characteristic results from cooperative binding of a sequential arrangement in the peptide of at least two binding modules, wherein each module is able to independently bind either streptavidin

or a streptavidin mutein. In this way, the multiple modules cooperate under non-competitive conditions in a synergistic manner to cause tight binding of the peptide to the substrate.

In the Schmidt declaration, comparative data is presented concerning binding to a streptavidin matrix of a fusion protein having a sequential arrangement of two peptide tags versus a single peptide tag. Under simple washing of the column with a buffer, the fusion peptide having a single tag begins to wash through the column, while the sequential tags bind more effectively. Under competitive elution conditions, however, each module of the sequential tags has to compete independently with binding of a free substrate binding molecule (or mimic thereof), *e.g.* biotin or biotin derivatives, thereby resulting in the effect that elution of the whole peptide that comprises the two sequentially binding modules is almost as fast as for a single module alone.

As described in the present specification (*e.g.*, in pages 3-8 and Example 3), the use of this sequential arrangement provides a surprising and substantial improvement in purification yield relative to the use of a single affinity tag under conservation of mild conditions for efficient elution by competitive displacement. This improvement in yield under preservation of mild conditions for the whole process has unquestionable practical benefit in the purification of recombinantly expressed proteins which often are labile and prone to denaturation under non-physiological conditions.

Furthermore, the surprising nature of this discovery is apparent from Szostak *et al.*, U.S. Patent No. 6,841,359, previously cited by the Examiner in the present case. In column 15, lines 63-66, the '359 patent reports that the presence of two HPQ (His–Pro–Gln) motifs does not confer high affinity binding. Likewise, column 10, lines 12-24 reports that binding to streptavidin is actually conferred by the entirety of a 38 residue peptide. In contrast, the present invention demonstrates that a much simpler sequential arrangement of binding motifs can provide sufficient binding affinity while maintaining the ability of the streptavidin binding peptides to be readily eluted under competitive mild conditions, and thus can provide improved purification yields.

C. No *prima facie* case of obviousness has been established

To conclude, Applicants respectfully submit that the secondary Srisawat *et al.* publication is not citable as prior art, and so no *prima facie* case of obviousness has been established. Moreover, the understanding and application of the secondary Srisawat *et al.* publication is in error. No motivation exists for combining the cited references is provided, and even if combined, the combination does not arrive at the claimed invention. Finally, the substantial evidence of nonobviousness already of record in the present case is sufficient to overcome any *prima facie* case that might be established.

In view of the foregoing, Applicants request that the rejection be reconsidered and withdrawn.

CONCLUSION

For the reasons set forth herein, Applicant respectfully submits that claims 16, 17, 32–34, 36, 37, 42–45, and 47-51 are in condition for allowance. Applicants respectfully request that the Examiner reconsider and withdraw the grounds for rejection set forth in the Office Action.

If the Examiner would like to discuss any of the issues raised in the Office Action, Applicant's representative can be reached at (619) 203-3186.

Respectfully submitted,

/Michael A. Whittaker/

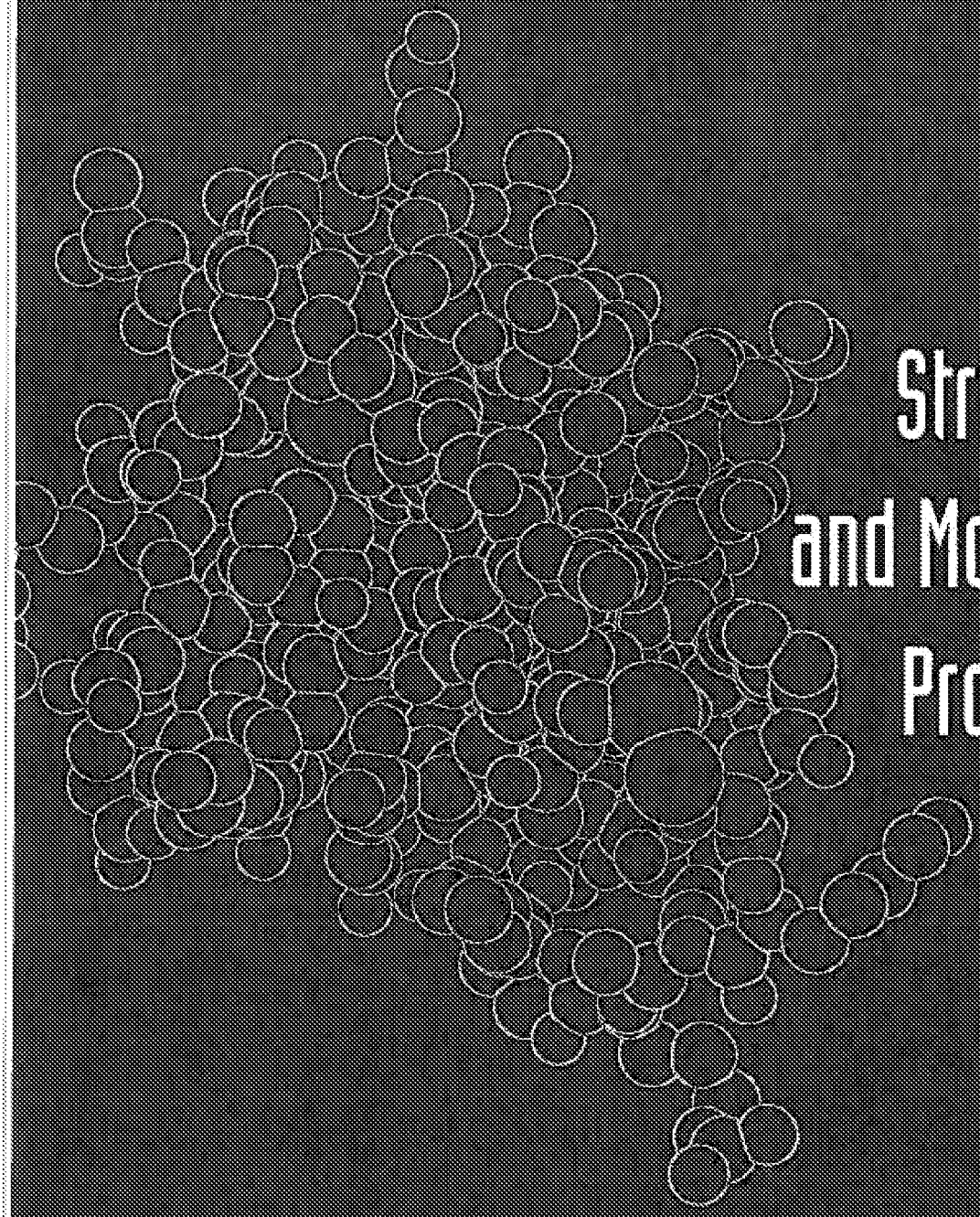
Date: December 2, 2010 By: _____

BIOTECHNOLOGY LAW GROUP
12707 High Bluff Drive
Suite 200
San Diego, CA 92103
Customer Number 35938

Michael A. Whittaker
Reg. No. 46,230
Tel. No. (619) 203-3186
Fax No. (858) 683-0390

SECOND EDITION

PROTEINS



Structures and Molecular Properties

THOMAS E. CREIGHTON

PROTEINS

Structures and Molecular Properties

Second Edition

Thomas E. Creighton
European Molecular Biology Laboratory
Heidelberg, Germany

W. H. Freeman and Company • New York

Cover image provided by Martin Noble.

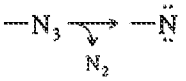
Library of Congress Cataloging-in-Publication Data
Creighton, Thomas E., 1940–
Proteins : structures and molecular properties / Thomas E. Creighton. — 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-7167-2317-4.
1. Proteins — Structure. 2. Proteins — Chemistry. I. Title.
QP551.C737 1993
574.19'245 — dc20 92-6664
CIP

Printed in the United States of America

1 2 3 4 5 6 7 8 9 0 KP 9 9 8 7 6 5 4 3

relaxed by attaching the reactive group to the ligand by a flexible "arm" that can come into contact with a number of groups on the protein that are within its radius. The price for this greater scope for reaction is that the effective concentrations of the protein and ligand groups are correspondingly lower, so that the reaction with the bound ligand is not so favored.

Bifunctional reagents have two reactive groups. One group can be used to react with the ligand, to generate the affinity label. This is then added to the protein, and the second group is encouraged to react with nearby portions of the protein. The reactivities of the two groups must be controlled so that each reacts only when required. Photoactivated groups are especially useful for the second step because they are reactive only in the presence of light. Also, they are then highly reactive with a variety of groups, making it likely that reaction with the protein will occur. Most commonly used are azide groups, which are totally inert until activated to the nitrene:



azide nitrene (8.1)

Nitrenes are reactive even toward methyl and methylene (—CH_2—) groups.

Bifunctional reagents with the same reactive group at both ends can be used to cross-link the protein and the ligand, which is very useful when the ligand is large or is also a protein, as in complicated macromolecular structures such as ribosomes, viruses, chromatin, and enzyme complexes. Whether or not the components of a complex are in close proximity can often be inferred by whether or not bifunctional reagents will cross-link them. When the components are polypeptide chains, those that have been cross-linked can usually be determined most readily by SDS electrophoresis (Sec. 1.5.3).

The versatility of the cross-linking approach lies in the ability to vary the length of the cross-linking group to serve as a molecular "ruler" for measuring the distances between two molecules in a complex. The nature of the bifunctional reagent can also be varied. For example, membrane proteins that are in proximity at the surface of the membrane can be cross-linked with a polar reagent that is confined to the aqueous solvent; in contrast, nonpolar reagents that permeate the membrane can cross-link protein molecules that interact in the membrane. Cross-linking with bifunctional reagents has been most useful with very large complexes, composed of numerous molecules, too large or too nonsymmetric to be studied crystallographically.

References

Chemical cross-linking: reagents and problems in studies of membrane structure. K. Peters and F. M. Richards. *Ann. Rev. Biochem.* 46:523–551 (1977).

Photoaffinity labelling of biological systems. V. Chowdhry and F. H. Westheimer. *Ann. Rev. Biochem.* 48:293–325 (1979).

Chemical cross-linking in biology. M. Das and C. F. Fox. *Ann. Rev. Biophys. Bioeng.* 8:165–193 (1979).

Mapping of contact areas in protein–nucleic acid and protein–protein complexes by differential chemical modification. H. R. Bosshard. *Methods Biochem. Anal.* 25:273–301 (1979).

Reaction of (bromoacetamido)nucleoside affinity labels with ribonuclease A: evidence for steric control of reaction specificity and alkylation rate. C. F. Hummel et al. *Biochemistry* 26:135–146 (1987).

Site-directed cross-linking. Establishing the dimeric structure of the aspartate receptor of bacterial chemotaxis. D. L. Milligan and D. E. Koshland, Jr. *J. Biol. Chem.* 263:6268–6275 (1988).

8.1.4 General Properties of Ligand Binding Interactions

In most cases, there is a single unique binding site for a particular ligand on each molecule of a polypeptide chain. If the polypeptide chain has internal symmetry, however, as when it has arisen by gene duplication (Sec. 3.4), each of the structural units may have a binding site. For example, gene-duplicated ferredoxins contain two similar iron–sulfur complexes related by the twofold symmetry of the molecule (see Fig. 6.36).

One structural domain can bind more than one ligand at separate binding sites, but it is unusual for there to be more than two or three such binding sites on any one domain. A protein that binds a number of different ligands often binds them on separate domains. These domains are often designated by their binding properties; an example is in Figure 8.3.

There are a few spectacular exceptions to the generalization of one binding site for a particular ligand per protein structural unit: Cytochrome c_3, a single polypeptide chain of only 118 residues, binds four identical heme groups in different environments. Bacteriochlorophyll protein from a green photosynthetic bacterium binds seven chlorophyll molecules, each in a different position within a "string bag" of a 15-stranded β-sheet closed to form a flattened barrel.

The earlier description of protein structures (Chap. 6) noted their tendency to be spherical. Binding sites for

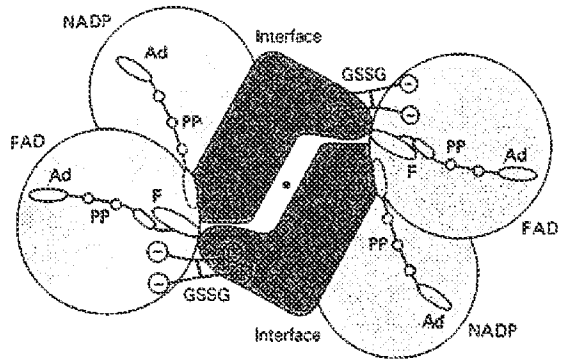


FIGURE 8.3
Schematic diagram of the domain structure of the dimeric enzyme glutathione reductase. The twofold axis relating the two polypeptide chains is indicated by the dot in the center. The two interface domains bind to each other and determine the dimeric structure. In the NADP-binding domain, the adenine (Ad), ribose (circles), diphosphate (PP), and nicotinamide (N) moieties are indicated; only the nicotinamide portion is completely buried. In the FAD-binding domain, the coenzyme is indicated similarly, except that F is the flavin moiety; only the adenine group is exposed to the solvent. The substrate, oxidized glutathione (GSSG), binds between the two subunits. (Adapted from G. E. Schulz et al., *Nature* 273:120–124, 1978.)

ligands are often sizable depressions on the surface and represent the greatest exceptions to this generality. It is probably more correct to state that the complex of a protein and its natural ligands tends to be spherical and that the interacting surface between them tends to be maximized. A very small ligand does not perturb the dimensions of a protein much and tends to be bound inside a relatively spherical protein molecule. Larger ligands tend to bind in depressions on the surface, where they can associate and dissociate. Those ligands that do not need to dissociate often or at all, such as hemes and some other prosthetic groups, tend to be bound deep in the protein interior and to be integral parts of the protein structure. Long, linear ligands such as polysaccharides tend to be bound in clefts on the surface. If the protein and its ligand are of similar size (e.g., two associating protein molecules), their interface tends to be flat and large. With very large ligands, such as nucleic acids, the protein tends to bind to depressions on the surface of the ligand.

Interactions between proteins and ligands demonstrate both steric and physical complementarity between the two. These interactions follow structural rules similar to those in the proteins themselves; indeed, many interfaces between two interacting protein molecules are not basically different from the interior of either protein. The interface between protein and ligand is usually as closely packed as is the protein interior. All polar groups in the interface are paired in hydrogen bonds, and electrostatic charges are generally neutralized. Hydrogen bonds are especially prominent in pairing polar groups, and water molecules frequently act as intermediaries. Ionic interactions are generally not buried in the interface unless it is necessary for charged groups on the ligand to be buried upon association; they then tend to be neutralized by groups of opposite charge. In other cases, however, there is no neutralizing charge, and the charge on the ligand appears to be effectively "solvated" by multiple hydrogen bonds to other groups of the protein (Fig. 8.4). Such hydrogen-bonding groups of the protein probably have fewer degrees of freedom than water molecules do, and they may offer a more stable solvation shell for the charged groups. The hydrogen-bonding groups are usually part of hydrogen-bond arrays that lead to the solvent or to groups with the opposite charge; the latter are in some

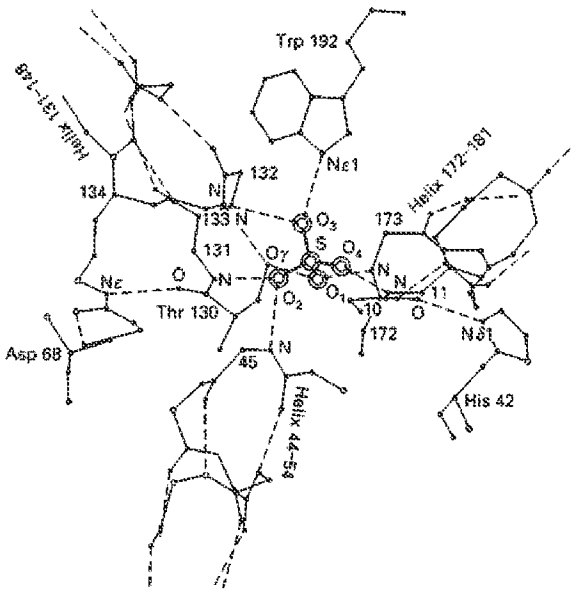


FIGURE 8.4
Binding of the SO_4^{2-} dianion to the sulfate-binding protein involved in bacterial active transport. The SO_4^{2-} molecule (depicted with double-circle atoms) is inaccessible to solvent, and there are no positively charged groups nearby. The crystal structure is consistent with there being seven hydrogen bonds between the ligand and the protein; all hydrogen bonds in the vicinity of the anion are indicated by dashed lines. (From F. A. Quiocho et al., *Nature* 329:561–564, 1987.)

Table 8.1 Conformational Changes Observed upon Ligand Binding to Some Nonallosteric Proteins

Protein	Ligand	Average Relative Movement (Å)	
		All atoms	Main chain
Trypsin	Bovine pancreatic trypsin inhibitor (BPTI)		0.26[a]
Trypsinogen	BPTI		0.28[b]
Lysozyme	Gd^{3+}	0.48[c]	
Myoglobin	O_2	0.92	0.61[d]
Concanavalin A	$Ca^{2+} + Mn^{2+}$	1.1	1.0[e]
Glyceraldehyde-3-phosphate dehydrogenase	NAD	1.2	1.2[f]
Carboxypeptidase A	Protein inhibitor		0.42[g]
Streptomyces griseus protease A	Ac-Pro-Ala-Pro-Phe-OH	0.11	0.10[h]
	Ac-Pro-Ala-Pro-Tyr-OH	0.10	0.09
	Ac-Pro-Ala-Pro-Phe-H	0.14	0.13

[a] From W. Bode and P. Schwyger, *J. Mol. Biol.* 98:693–717 (1975).

[b] From W. Bode et al., *J. Mol. Biol.* 118:99–112 (1978).

[c] From S. J. Perkins et al., *Biochem. J.* 173:607–616 (1978).

[d] From S. E. V. Phillips, *J. Mol. Biol.* 142:531–554 (1980).

[e] From M. Shoham et al., *J. Mol. Biol.* 131:137–155 (1979).

[f] From M. R. N. Murthy et al., *J. Mol. Biol.* 138:859–872 (1980).

[g] From D. C. Rees and W. N. Lipscomb, *Proc. Natl. Acad. Sci. USA* 78:5455–5459 (1981).

[h] M. N. G. James et al., *J. Mol. Biol.* 144:43–88, (1980).

cases the ends of α-helices, which have partial charge as a result of the helix dipole (Sec. 5.3.1). The hydrogen-bond arrays seem to have the role of effectively dispersing the formal buried charge.

The structure of a protein domain generally does not change substantially when it binds a ligand; exceptions are usually of functional importance. Small movements of atoms of the protein do occur in every case (Table 8.1), but they are often comparable to the experimental errors in crystallographic structure analysis. The most extreme changes in domains generally involve movements of flexible loops on the protein surface. On the other hand, some small adjustments are probably important in general to permit rapid rates of association and dissociation; totally rigid complex structures in which atoms interlock and interdigitate would be unlikely to be able to come together readily.

The analogy is often made with a key fitting into a lock. Although the rigidity implied by this analogy is too extreme for proteins, it conveys the correct message that a defined protein structure is probably necessary for specificity in ligand binding (Fig. 8.5). A very malleable protein would adopt its shape to match that of many ligands and would bind many of them with similar affinities. The difference in the affinities for two ligands is limited by the energy required to distort the normal conformation that is complementary to the high-affinity ligand to a conformation that is complementary to the low-affinity ligand. In the few known cases, binding of ligands with low affinity to a protein is not observed to produce large changes in conformation of the protein. The low affinities of such ligands primarily reflect their noncomplementarity to the preexisting binding site. The difference in binding energies for low- and high-affinity ligands is not sufficient to distort the protein binding site to make it complementary to the low-affinity ligand (Fig. 8.5).

Substantial changes in protein structure upon ligand binding are most frequently limited to motions of rigid domains or subunits relative to each other. Many ligands bind between domains that move together to engulf the ligand. This may help to maximize the inter-

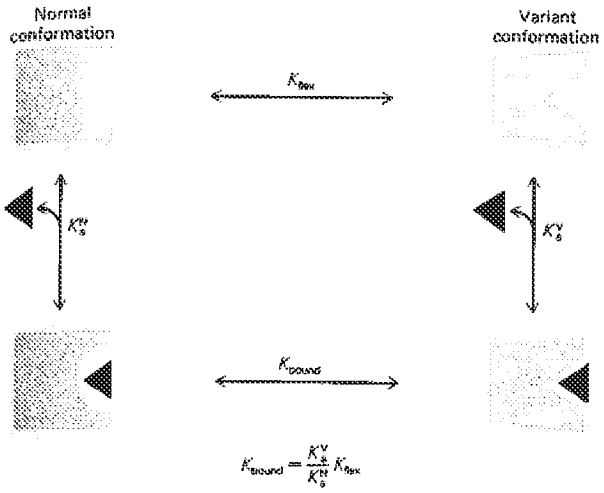


FIGURE 8.5
Estimation of protein flexibility by the conformational distortions produced by binding a noncomplementary ligand. The normal conformation of a protein is imagined not to be complementary to a ligand (designated by the triangle) and consequently has poor affinity for that ligand. A variant conformation of the protein resulting from its flexibility might be complementary to the ligand and bind it tightly, so $K_a^V \gg K_a^N$, but this conformation would not normally be populated substantially (i.e., $K_{flex} \ll 1$). When the ligand is bound, however, the variant conformation should be more stable by the ratio of the two binding affinities, so it might be populated in the liganded state. For example, if $K_{flex} = 10^{-3}$ and the variant conformation binds the ligand 10^4 times more tightly than the normal conformation, K_{bound} will be 10, so the variant conformation should be present in the complex 91% of the time. On the other hand, if no substantial changes in the protein are observed upon binding a poor ligand, the protein must be relatively inflexible, so K_{flex} is extremely small.

actions between the protein and the ligand and to minimize interactions with other components of the solvent, while permitting the ligand to associate and to dissociate. The other role of domain or subunit movements upon ligand binding is to produce functional alterations at other sites on the protein, as in allosteric proteins (Sec. 8.4). Binding in this case is frequently regulated by changes elsewhere in the protein. Binding sites between domains or subunits seem to be dynamic locations, probably owing to the much greater flexibility of the protein's quaternary structure than of its tertiary structure.

In a few cases, binding of a ligand does produce substantial changes in the structure of protein domains. An example is the effects of Ca^{2+}-binding by regulatory proteins such as calmodulin (Sec. 8.3.4.a).

From these general principles, it is often possible to guess correctly the structure of a protein–ligand complex if the structures of the two components are known. Nevertheless, do not underestimate the difficulty of fitting together two molecules of known structure to make the most stable complex. There are an enormous number of ways that two molecules can associate, especially if at least one is the size of a protein, and only one of the ways is likely to be the correct one.

References

Structural and functional aspects of domain motions in proteins. W. S. Bennett and R. Huber. *Crit. Rev. Biochem.* 15:291–384 (1984).

Computer studies of interactions between macromolecules. S. J. Wodak et al. *Prog. Biophys. Mol. Biol.* 49:29–63 (1987).

Stabilization of charges on isolated ionic groups sequestered in proteins by polarized peptide units. F. A. Quiocho et al. *Nature* 329:561–564 (1987).

Substrate specificity and affinity of a protein modulated by bound water molecules. F. A. Quiocho et al. *Nature* 340:404–407 (1989).

High specificity of a phosphate transport protein determined by hydrogen bonds. H. Luecke and F. A. Quiocho. *Nature* 347:402–406 (1990).

The structure of protein–protein recognition sites. J. Janin and C. Chothia. *J. Biol. Chem.* 265:16027–16030 (1990).

8.2 Energetics and Dynamics of Binding

A fundamental aspect of the interaction of a protein with a ligand is the affinity of the two for each other, which is a measure of the overall free energy of the interaction. The magnitude of the affinity determines whether a particular interaction is relevant under a given set of conditions. The observed affinities of proteins for ligands vary enormously, ranging from very high values, for which dissociation is immeasurably small, to very low values, for which the concentration of free ligand required for a significant degree of binding is so great as to cast doubt on its relevance. Whether or not any particular affinity of a protein for a ligand is significant depends on the concentration of the ligand that the protein is likely to encounter; no other generalizations are possible.

If the affinity is very high, the protein is likely to be found and isolated as the complex; if such a ligand is relatively small, it is designated a **prosthetic group.** Examples are the heme groups of the globins and cy-

tochromes, some coenzymes that bind tightly to enzymes, and metal ions that are integral parts of the protein structure. With lower affinities, ligands that are originally bound to a protein are likely to be lost during purification, unless they are added to the protein solution.

8.2.1 *Binding Affinities*

The affinity between a protein P and a ligand A is measured by the **association constant** K_a for the binding reaction at equilibrium:

$$P + A \underset{}{\overset{K_a}{\rightleftharpoons}} P \cdot A \tag{8.2}$$

$$K_a = \frac{[P \cdot A]}{[P][A]} \tag{8.3}$$

All species are presumed to be present at sufficiently low concentrations for thermodynamic ideality to apply; if not, activities rather than concentrations must be measured. K_a is a constant under a given set of conditions and is measured experimentally by the dependence of binding on the free ligand concentration. Several commonly used graphic methods of analyzing binding data are illustrated in Figure 8.6.

The ratio of bound to free protein should be, according to Equation (8.3), directly proportional to the free-ligand concentration:

$$\frac{[P \cdot A]}{[P]} = K_a[A] \tag{8.4}$$

An experimentally more useful measure of binding is the fraction y of protein molecules with bound ligand:

$$y = \frac{[P \cdot A]}{[P] + [P \cdot A]} = \frac{K_a[A]}{1 + K_a[A]} \tag{8.5}$$

The greater the value of K_a, the greater the affinity. The value of K_a has units of (concentration)$^{-1}$, however, and it is often intuitively easier to consider the **dissociation constant** K_d, which is simply the reciprocal of K_a and has units of concentration. With concentrations of free ligand below K_d, little binding to the protein occurs. With a concentration equal to K_d, half the protein molecules have bound ligand. An occupancy of 90% requires a nine times greater concentration of free ligand, whereas 99% occupancy requires that the concentration be 99 times K_d. Binding equilibria are simplest when the ligand is present at a concentration much greater than that of the protein binding sites. Uptake of the ligand by the protein does not then significantly alter the concentration of free ligand.

Specific binding by a protein of one ligand, and not another, depends on their relative affinities, their concentrations, and whether they bind at the same site. If two ligands are present at a concentration of 10^{-5} M but have different values of K_d—say, 10^{-3} M and 10^{-6} M—only the ligand with the lower K_d is bound significantly. If both are present at much higher concentrations—say, 10^{-2} M—both are bound to the protein to the maximum extent if they bind at separate sites. In this case, the higher affinity of one ligand is almost immaterial. If the two ligands compete for the same site, however,

$$P \cdot A \overset{A}{\underset{K_d^A}{\rightleftharpoons}} P \overset{B}{\underset{K_d^B}{\rightleftharpoons}} P \cdot B \tag{8.6}$$

$$[P] = \frac{[P \cdot A]K_d^A}{[A]} = \frac{[P \cdot B]K_d^B}{[B]} \tag{8.7}$$

$$\frac{[P \cdot A]}{[P \cdot B]} = \frac{K_d^B}{K_d^A}\frac{[A]}{[B]} \tag{8.8}$$

the ligand with the higher affinity is bound to a correspondingly greater extent when the ligands are present at the same concentration. Weaker affinity can always be overcome by a higher concentration of that ligand, however, so binding affinities should always be considered relative to the concentration of the ligand.

The energetics of binding are often expressed by the Gibbs free energy of binding, ΔG_{bind}:

$$\Delta G_{bind} = -RT \ln K_a = RT \ln K_d \tag{8.9}$$

It must be kept in mind, however, that K_a and K_d have units of concentration and that the value of ΔG_{bind} depends on which units are used (i.e., the standard state). If the units are moles per liter, the standard state is 1 M, and the calculated value of ΔG_{bind} applies only under the rather arbitrary situation when the concentration of free ligand is 1 M. In many instances a "unitary" free energy of binding is used as a measure of the intrinsic affinity; this is the free energy of binding that would occur with ligand at a hypothetical concentration of 55 M, the normal concentration of water. This parameter is not of any special significance, however, except when the ligand is water, and it does not represent the free energy of interaction that would occur in a unimolecular interaction (see Sec. 8.2.2).

The energetics of binding are defined more explicitly as the difference in free energies of the free and liganded protein, ΔG_b:

$$\Delta G_b = -RT \ln (K_a[A]) = -RT \ln \left(\frac{[A]}{K_d}\right) \tag{8.10}$$

In this case, the concentration of free ligand must be specified. In a similar way, the enthalpy and entropy of

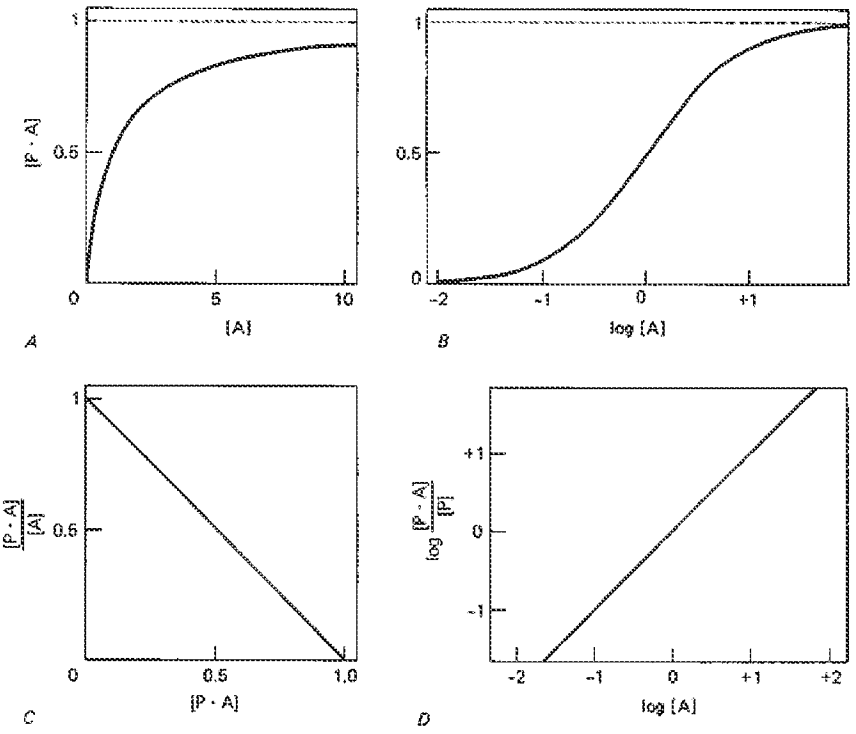


FIGURE 8.6
Some common methods of plotting binding data, using theoretical curves for the simple binding reaction $P + A \rightleftharpoons P \cdot A$. The concentration of free ligand is expressed relative to its dissociation constant, which is that concentration of free ligand that gives half-maximal binding. The concentrations of free and liganded forms of the proteins are given relative to the concentration of total protein. *A:* The normal hyperbolic relationship between binding and free-ligand concentration, demonstrating that a free-ligand concentration 9-fold greater than its dissociation constant produces only 90% of maximal binding (indicated by the dashed line); a 99-fold greater concentration is required for 99% saturation. *B:* A logarithmic scale emphasizes the wide range of free-ligand concentrations required for a complete binding curve. *C:* Scatchard plot. The negative slope gives the value of the association constant (the reciprocal of the dissociation constant). The horizontal intercept gives the extrapolated extent of the maximal binding. *D:* Hill plot. An accurate value for the maximum binding is required for this plot because both the liganded and the free protein concentrations are required. The value of the dissociation constant is given by the value of the free-ligand concentration where the vertical axis is zero (i.e., at half-maximal binding). This plot is used primarily for analyzing cooperative binding (see Fig. 8.24C).

binding are defined by the temperature dependence of the binding affinity.

A general consequence of ligand binding is that the protein is stabilized against unfolding and is less flexible. Neither of these observations need imply that the ligand has altered the structure of the protein. Instead, they are simply a consequence of the ligand binding more tightly to the fully folded conformation (N) than to the fully unfolded state (U) and any distorted or partially unfolded forms that result from flexibility of the structure. This can be illustrated very simply for the case of unfolding of the protein when the ligand L binds solely to the folded state N:

$$N \cdot L \overset{L}{\underset{K_d^L}{\uparrow}} N \overset{K_u}{\longleftrightarrow} U \qquad (8.11)$$

$$K_{app} = \frac{[N \cdot L] + [N]}{[U]} = K_u \left(1 + \frac{[L]}{K_d^L}\right) \qquad (8.12)$$

The protein is stabilized against unfolding by the presence of the ligand. Even at very high ligand concentra-

tions, above those at which the folded protein is fully saturated, the apparent stability of the protein is increased in proportion to the concentration of free ligand.

Ligand binding is simple in dilute solutions, but proteins often function in extremely concentrated aqueous solutions, as in the cytosol. For example, the interior of the red blood cell is about 35% hemoglobin by weight. Such solutions are very nonideal. The pertinent equilibria must be expressed in terms of the thermodynamic activities of the protein and of the ligand, which can be very different from their concentrations. Even though a particular protein might not be present in high concentration, the presence of molecules other than water in the environment can lead to substantial excluded volume effects (see Fig. 7.2). Added molecules favor any conformational or binding reaction that leads to a more spherical shape of a protein molecule, with less surface area exposed to solvent (Sec. 7.1.1). Consequently, binding of a ligand to a protein is often considerably greater in a concentrated solution than might be expected. It is possible that most proteins in the cytosol usually exist bound to each other, to membranes, to cytoskeleton, or to some other organized structure.

References

The meaning of Scatchard and Hill plots. F. W. Dahlquist. *Methods Enzymol* 48:270–299 (1978).

Protein affinities for small molecules: conceptions and misconceptions. I. M. Klotz and D. L. Hunston. *Arch. Biochem. Biophys.* 193:314–328 (1979).

Excluded volume as a determinant of macromolecular structure and reactivity. A. P. Minton. *Biopolymers* 20:2093–2120 (1981).

Calorimetric approaches to protein–ligand interactions. H. J. Hinz. *Topics Mol. Pharmacol.* 71–122 (1983).

Characterization of binding equilibrium data by a variety of fitted isotherms. R. Broderson et al. *Eur. J. Biochem.* 169:487–495 (1987).

An accurate method for determination of receptor–ligand and enzyme–inhibitor dissociation constants from displacement curves. A. Horovitz and A. Levitzki. *Proc. Natl. Acad. Sci. USA* 84:6654–6658 (1987).

Study of strong to ultratight protein interactions using differential scanning calorimetry. J. F. Brandts and L. N. Lin. *Biochemistry* 29:6927–6940 (1990).

8.2.2 *Accounting for Relative Affinities*

Can we account for the particular affinity of a protein for a given ligand? Do the structures of the two give any insight into why the affinity is high or low? A qualified yes is permissible in answer to these questions in certain cases. Generally, however, only qualitative conclusions are feasible. Yet it should be possible eventually to account for, or even to predict, binding affinities for ligands and to design ligands or proteins with useful binding properties. The implications are immense for chemotherapy and for drug design.

It is currently impossible to rationalize the values of K_a or K_d of any protein for any ligand, even when the structures of the complex and of the free components are known. The practical difficulties in rationalizing ligand affinities arise from our poor understanding of the energetics of protein structure (Chap. 7) and of the strengths of the basic interactions in aqueous solvent (Chap. 4). The observed affinity depends on the relative free energies of the complex and of the components. This includes not only the interactions between the two in the complex but also any changes in their average conformations and their flexibilities produced by complex formation, any differences in their various interactions with solvent, the loss of translational and rotational freedom of each component, plus the displacement of solvent and any other ligands present in the binding sites before formation of the complex. Many of these factors compensate each other, and the net observed effect is a small difference between several terms of large and uncertain magnitude. It is not yet possible to calculate these quantities accurately enough to predict the values of K_a and K_d.

The relative affinities of two related ligands for the same protein (e.g., A and B), or of two closely related proteins for the same ligand, are more easily analyzed because many of the factors are the same in the two cases; differences in affinities can often be related to just one or two factors. Also, the ratio of their affinities is dimensionless:

$$\Delta(\Delta G^\circ)_{A-B} = -RT \ln \frac{K_a^A}{K_a^B} = +RT \ln \frac{K_d^A}{K_d^B} \quad (8.13)$$

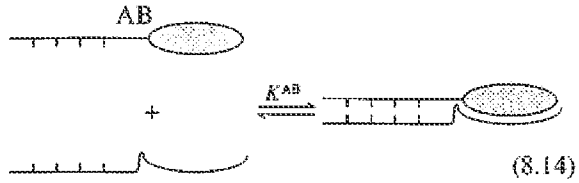
There is then no complication in defining standard states. The most successful method for analyzing differences in binding interactions between closely related ligands or proteins is the free-energy perturbation method (Sec. 7.4.4.b) for simulating the effects of differences in chemical structure of the ligand or the protein, where the free energy is calculated in both the complex and the free molecule as the group that differs is gradually "mutated" during the calculation.

Interactions between a protein and a ligand always involve a substantial number of groups. The general approach to understanding the observed affinity has been to dissect it into the contributions of each group by measuring the effect on the affinity of removing them

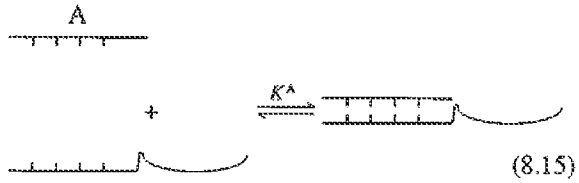
individually. Varying the ligand has traditionally been easiest, but even closely related ligands are occasionally observed to bind to the same protein in very different ways; in this case, a comparison of their measured binding affinities is largely meaningless. It is now more acceptable to use the same ligand but to vary the protein, using site-directed mutagenesis (Sec. 2.2), because the structures of the variant proteins tend to remain more constant. It is still advisable, however, to determine that the ligand binds in the same way to the variant proteins.

Given a series of binding affinities of related ligands for the same protein or of variant proteins for the same ligand, how is the binding energy dissected? It might be thought that the total binding energy (Eq. 8.9) is simply the sum of the contributions of each group, but it is not that straightforward. This can be illustrated in a manner first presented by Jencks.

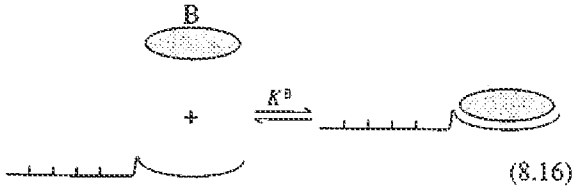
Consider a ligand composed of two parts A and B; A might be capable of hydrogen bonding, and B might be hydrophobic. The affinity of ligand AB is compared with the affinities of A and B separately:



(8.14)



(8.15)



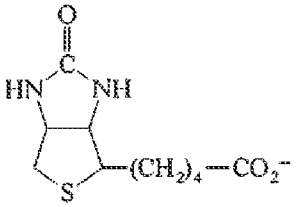
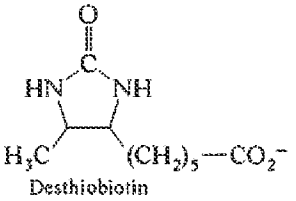
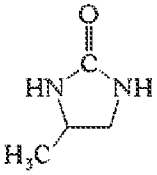
(8.16)

where the equilibrium constants K^{AB}, K^{A}, and K^{B} are for either association or dissociation. In general, there is no simple relationship among these constants, and the classical binding energies calculated from them using Equation (8.9) are generally not additive:

$$-RT \ln K_a^{AB} \neq -RT \ln K_a^{A} - RT \ln K_a^{B} \quad (8.17)$$

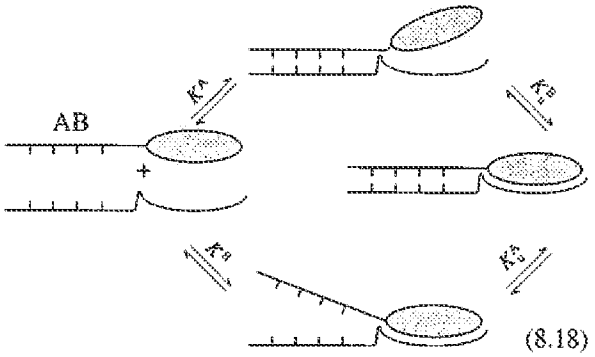
even if the standard state is taken as 55 M so that unitary binding energies are calculated (Sec. 8.2.1). This nonadditivity is illustrated by the binding of biotin and some derivatives to the protein avidin (Table 8.2).

Table 8.2 Binding of Biotin Derivatives to Avidin

Derivative	Dissociation constant (M)	Free-energy contribution to binding (kcal/mol)
 Biotin	1.3×10^{-15}	
 Desthiobiotin	5×10^{-13}	
	3.4×10^{-5}	−13.3
$CH_3—(CH_2)_4—CO_2^-$	3×10^{-5}	−10.7

Data from N. M. Green, *Adv. Protein Chem.* 29:85–133 (1975).

The reason for the nonadditive nature of specific binding affinities becomes clear if the binding of AB is dissected into steps:



(8.18)

The first step in the binding of the entire ligand should be analogous to the binding of each part when present

alone (Eqs. 8.15 and 8.16), so the first two steps are assigned the bimolecular binding constants of each part.

The second step, binding of the second part of the ligand, however, is now a unimolecular step, rather than bimolecular. Consequently, the bimolecular binding constants K^A and K^B do not apply to the second step; instead, these second steps are assigned the unimolecular equilibrium constants K_u^A and K_u^B. The values of these two constants are not independent because they are linked functions:

$$K_a^{AB} = K_a^A K_u^B = K_a^B K_u^A \tag{8.19}$$

From Equation (8.19), it is apparent that the constants K_u^A and K_u^B are the ratios of the binding constants of ligands with and without each of the respective moieties:

$$K_u^A = \frac{K_a^{AB}}{K_a^B} \qquad K_u^B = \frac{K_a^{AB}}{K_a^A} \tag{8.20}$$

The contribution to the free energy of binding of each moiety can then be calculated:

$$\Delta\Delta G_f^A = -RT \ln K_u^A = -RT \ln \frac{K_a^{AB}}{K_a^B} \tag{8.21}$$

$$\Delta\Delta G_f^B = -RT \ln K_u^B = -RT \ln \frac{K_a^{AB}}{K_a^A} \tag{8.22}$$

For example, the free-energy contribution to binding of the sulfur atom of biotin can be estimated from the relative affinities of the first two compounds of Table 8.2 to be −3.5 kcal/mol:

$$\Delta\Delta G_f^S = -RT \ln \frac{5 \times 10^{-13}\ M}{1.3 \times 10^{-15}\ M} = -3.5 \text{ kcal/mol} \tag{8.23}$$

Similarly, the contributions of the remaining five-membered ring and of the acidic hydrocarbon group can be estimated to be −13.3 and −10.7 kcal/mol, respectively (Table 8.2).

The incremental binding energy contributions calculated in this way give a measure of the increased affinity caused by the presence of each group of the ligand. Their values depend critically on the relationship between the two parts of the ligand during binding, that is, on the effective concentration of the second part of the ligand when the first is bound, which can be designated as [A/B].

$$K_u^A = K_a^A\,[\text{A/B}] = \frac{[\text{A/B}]}{K_d^A} \tag{8.24}$$

$$K_u^B = K_a^B\,[\text{A/B}] = \frac{[\text{A/B}]}{K_d^B} \tag{8.25}$$

The same effective concentration applies to both parts because these are linked functions (see Eq. 8.19).

The effective concentration of either part in an intermediate complex can conceivably be zero, when it is kept away from the binding site and so provides no contribution to binding. Or, it might have a value of up to 10^{10} M (see Table 4.11), when parts A and B of the ligand are always in optimal orientation for simultaneous binding to a perfectly complementary binding site. The large values of effective concentrations in intramolecular reactions result from the entropic effect of the covalent linkage of the two parts. A ligand must lose a substantial amount of translational and rotational entropy upon binding; this is one of the factors determining the values of both K^A and K^B. In the case of ligand AB, however, at least some of this entropy is lost when the first part is bound; the second part of the ligand is then fixed to some extent and need not lose as much entropy upon completion of the binding as if that part were binding by itself. Consequently, the greater the rigidity between the two parts of a ligand, the greater the entropic contribution to the effective concentration is likely to be. This entropic contribution is the primary reason that the binding contributions of individual parts of a ligand do not add to give the observed affinity. For example, the contributions to the binding of biotin by the sulfur atom, the five-membered ring, and the acidic hydrocarbon (Table 8.2) total −27.5 kcal/mol. In contrast, the free energy of binding that would be calculated from Equation (8.9) is only −20.2 kcal/mol. The difference between these values reflects primarily the greater entropy that must be lost when parts of a ligand bind as separate molecules relative to the entropy that must be lost when they bind as parts of the same molecule.

If neither the ligand nor the protein is strained by binding, very high effective concentrations and free-energy contributions to binding may be observed. For example, the data of Table 8.2 for the two halves of desthiobiotin imply that their effective concentrations in the hypothetical intermediate complex are 2×10^5 M because

$$[\text{A/B}] = \frac{K_a^{AB}}{K_a^A K_a^B} = \frac{K_d^A K_d^B}{K_d^{AB}} \tag{8.26}$$

With such high effective concentrations, ionic and hydrogen-bond interactions between ligand and protein may contribute substantially to binding, even in aqueous solution, where they must compete with intermolecular interactions between the solvent and the free protein and free ligand.

Because effective concentrations of the different parts of ligands are likely to vary substantially in different ligands and different binding situations, it is unrealistic to expect a constant contribution to binding of a hydrogen bond, a van der Waals interaction, and so on, in all ligand-binding interactions.

Table 8.3 Large Contributions to Ligand Affinities for Proteins

Group of ligand	Free-energy contribution to binding to protein[a] (kcal/mol)	Free energy of transfer from water to nonpolar liquid[b] (kcal/mol)
—CH_3	−2.0 to −3.9	−0.5
—CH_2CH_3	−6.5	−1.0
—CH—$(CH_3)_2$	−9.6	−1.5
—CH_2—CH_2—CH_2—CH_3	−7 to −8	−2.6
—SCH_3	−4.9	
—CH_2—CH_2—S—CH_3	−10 to −11	−1.3
—SH	−5.4 to −9.1	
—OH	−8	
—NH_2	−4.5	
—NH_3^+	−6.7	
—CO_2^-	−4.3	

[a] Contributions to binding were measured by the difference in affinities of ligands that differ only in the presence or absence of the indicated group; the free-energy contribution was calculated using Equation (8.21). Data from W. P. Jencks, *Proc. Natl. Acad. Sci. USA* 78:4046–4050 (1981); A. R. Fersht, *Proc. Roy. Soc. Lond. [Biol.]* 212:351–379 (1981).

[b] Data from Y. Nozaki and C. Tanford, *J. Biol. Chem.* 246:2211–2217 (1971).

Some examples of large incremental contributions to binding by various groups, measured by the relative affinities of ligands that differ only in that group, are tabulated in Table 8.3. The values for the nonpolar groups are considerably greater than their free energies of transfer from water to nonpolar liquids, which often is considered an analogous process. This discrepancy is further evidence that a protein — at least, its binding site — is not equivalent to an organic liquid. Instead, the folded protein has a higher concentration of atoms, and a binding site for nonpolar groups probably presents a more rigidly defined cavity with greater van der Waals interactions than is possible for a liquid. If part of the ligand or the protein normally involved in binding is missing, there might be a void at the interface between protein and ligand. Such a void could be filled by an isolated solvent molecule, or the protein and ligand could adapt to attain complementarity; but both are energetically costly. If a polar group normally involved in hydrogen bonding is deleted, its partner can be left in an energetically unfavorable situation without an alternative group to hydrogen-bond to. Consequently, interpreting such binding data in terms of individual interactions is not straightforward.

The data of Table 8.3 demonstrate that a protein can discriminate very effectively between its proper ligand and a ligand that lacks just one small part. Discrimination of ligands containing extra groups can be even more powerful because additional groups can interfere sterically with the complementarity between ligand and binding site.

Nevertheless, there are limits to the specificity of binding that is possible, set by the energetics of the interactions between groups. These limits are exceeded in some instances for which extreme specificity is necessary: for example, in the replication, transcription, and translation of genetic information. DNA replication occurs with an error frequency of only 10^{-10} even though the tautomerization of the nucleic acid bases, which will cause incorrect base-pairing, occurs with a frequency of 10^{-5}. Amino acids are also incorporated into proteins with considerably greater fidelity than expected, even from the data of Table 8.3. Much of that data comes from binding of amino acids to tRNA synthetases, which carry out the most crucial step of attaching the correct amino acid to the correct tRNA molecule. For example, how does a tRNA synthetase discriminate effectively against Gly when adding Ala to its tRNA, Val in the case of Ile, and Ser in the case of Thr? These pairs differ only by one —CH_2— group and might be expected (Table 8.3) to differ in affinity by only a factor of 10^2. The answer in this case is that the enzyme seems to check the amino acid twice, discriminating at the first binding step and then subjecting the selected

amino acid to a second check designed to detect the most likely fraudulent amino acid. Any caught by the second step are hydrolyzed from the tRNA and expelled (see Sec. 9.3.1.b). In such a *double-sieve* editing mechanism, the probability that an incorrect amino acid will be missed by both steps is the product of the two separate probabilities (e.g., $10^{-5} \times 10^{-5} = 10^{-10}$). In this way, biological specificities can be greatly enhanced over those possible with simple physical principles. Comparable multiple checks on specificity appear to be used in DNA replication. Such methods are used only when absolutely necessary, however, because there is a cost involved, in that a certain fraction of correct molecules are also removed at the subsequent recognition steps due to the intrinsic limitations on binding specificity.

References

Binding energy, specificity, and enzymic catalysis. W. P. Jencks. *Adv. Enzymol.* 43:219–410 (1975).

Entropy, binding energy, and enzymic catalysis. M. I. Page. *Angew. Chem. Int. Ed. Engl.* 16:449–459 (1977).

On the attribution and additivity of binding energies. W. P. Jencks. *Proc. Natl. Acad. Sci. USA* 78:4046–4050 (1981).

Effects of engineering complementary charged residues into the hydrophobic subunit interface of tyrosyl-tRNA synthetase. W. H. J. Ward et al. *Biochemistry* 26:4131–4138 (1987).

Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis. A. R. Fersht. *Biochemistry* 27:1577–1580 (1988).

The hydrogen bond in molecular recognition. A. R. Fersht. *Trends Biochem. Sci.* 12:301–304 (1987).

Computer modeling of the interactions of complex molecules. P. A. Kollman and K. M. Merz, Jr. *Acc. Chem. Res.* 23:246–252 (1990).

8.2.3 Rates of Binding and Dissociation

The rates of binding and of dissociation of a ligand from a protein are determined by the respective rate constants k_a and k_d:

$$\mathrm{P} + \mathrm{A} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{P} \cdot \mathrm{A} \qquad (8.27)$$

Their ratio gives the association constant:

$$K_a = \frac{k_a}{k_d} \qquad (8.28)$$

The rate constants for binding ligands to proteins vary considerably, depending on the sizes of both and on any conformational changes that must take place in each upon binding. Many small ligands are found to bind very rapidly, at rates approaching those expected for diffusion control, k_D. This expected rate can be estimated from the diffusion coefficients of the protein and ligand, D_P and D_A, respectively, treating them as small spherical molecules that must approach within a distance r_{PA} for binding to occur:

$$k_D = 4\pi N_A (D_P + D_A) r_{PA} \qquad (8.29)$$

where N_A is Avogadro's number. For molecules with typical diffusion coefficients under normal circumstances (see Table 7.2), values of k_D in the region of $10^9\ M^{-1}\ s^{-1}$ are expected. Larger molecules have smaller diffusion coefficients, but the value of k_D does not decrease accordingly because the value of r_{PA} is correspondingly larger. If the two molecules attract or repel each other at a distance, the term r_{PA} in Equation (8.29) should be replaced by a term containing the energy of interaction as a function of distance. For example, electrostatic interactions are significant over substantial distances and, when favorable, can increase rate constants for association to $10^{11}\ M^{-1}\ s^{-1}$.

Rates of binding that are observed to be lower than k_D imply either that the two molecules must be in defined orientations for productive binding to occur or that changes occur during binding to produce a multistep association reaction. Both phenomena are undoubtedly important with proteins. The binding sites on proteins usually bind ligands only in defined orientations, and they generally comprise only small fractions of the protein surface. Consequently, most encounters between ligand and protein would be expected to be unproductive, and association to be relatively slow, but there are exceptions. For example, cytochromes *c* are thought to transfer electrons to and from other proteins through only 0.6% of their surfaces, where the heme group is accessible (see Fig. 6.31), and only when the two proteins interact in very specific orientations. On this basis, the rate of their interaction would be expected to be lower than that for diffusion-controlled encounters by a factor of at least 1000, but it is not. The reason is thought to be that asymmetric distributions of charges on the proteins orient them so that they tend to approach each other rapidly in a productive manner. The charge distribution of horse cytochrome *c* indicates a large dipole moment of just over 300 Debye units, and the dipole axis passes through the presumed binding site. Electrostatic interactions have also been shown to guide charged ligands to their binding sites on other proteins.

Table 8.4 Rate Constants for Formation of Complexes of tRNA Synthetase (P) and tRNA[a]

$$P + tRNA \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} P \cdot tRNA \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} P * tRNA$$

		Initial Binding and Dissociation		Isomerization of Complex	
tRNA synthetase	tRNA	k_1 $(M \cdot s)^{-1}$	k_{-1} (s^{-1})	k_2 (s^{-1})	k_{-2} (s^{-1})
Yeast Ser[b]	Yeast Ser	2.7×10^8	220	760	330
Yeast Phe[c]	Yeast Phe	2×10^8	250	420	750
Yeast Phe[c]	*E. coli* Tyr	8×10^8	1600	—	—

[a] The source and amino acid specificity of each tRNA synthetase and tRNA are indicated. The difference between the two complexes P · tRNA and P * tRNA is not known.

[b] Data from D. Riesner et al., *Eur. J. Biochem.* 68:71–80 (1976).

[c] Data from G. Krauss et al., *Eur. J. Biochem.* 68:81–93 (1976).

Binding of a ligand to a protein probably occurs generally via diffusion-controlled formation of an unstable *encounter* complex, followed by its rearrangement to the final complex:

$$P + A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \{P \cdots A\} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} P \cdot A \qquad (8.30)$$

Such encounter complexes are usually not observed, and their structures are not known. The observed rate of binding is still proportional to the free-ligand concentration but is slower than diffusion-controlled k_1 because the encounter complex is unstable (i.e., $k_{-1} > k_1[A]$), and it dissociates more rapidly than it completes binding (i.e., $k_{-1} > k_2$). Such encounter complexes can be observed only if high ligand concentrations are used so that k_2 becomes rate-limiting. Their occurrence may be important for permitting the two reactants to find their appropriate orientations for tight binding, rather than relying on the two molecules being in exactly the correct orientation in the initial encounter.

Even if the observed rate of association is apparently diffusion-controlled, additional steps may follow the initial association. For example, the initial interaction between a tRNA molecule and a tRNA synthetase enzyme occurs at nearly the diffusion-controlled rate, about $10^8/(M \cdot s)$, but rearrangements occur subsequently on the millisecond time scale (Table 8.4). Very little is known about what occurs structurally during the course of binding of ligands to proteins. There is considerable scope for conformational changes in both protein and ligand in most cases, but techniques for following them have yet to be devised.

Diffusion-limited encounters occur between all molecules in a solution; therefore, stable and specific binding must be reflected primarily in slow rates of dissociation, k_d (Eq. 8.28). Energetically favorable rearrangements of the complex after initial association have the effect of decreasing the apparent rate of dissociation. For example, binding of a yeast tRNA synthetase to a bacterial tRNA is not productive and is weaker than with the homologous pairs, but the initial association is somewhat faster. The weaker binding results from a greater rate of dissociation of the initial complex and from the apparent absence of a second isomerization step (Table 8.4).

The analysis just presented has been addressed to solutions in which the molecules are free to diffuse in three dimensions. In this case, the rate of diffusion-limited association depends on the size of the target, but this is no longer true when diffusion is confined to a space of lower dimensions, such as the two-dimensional plane of a membrane. For example, the mean diffusion times τ to reach a small target of radius a in the middle of a space of radius R ($R \gg a$) are given by

$$\tau_3 = \left(\frac{R^2}{3D_3}\right)\left(\frac{R}{a}\right) \quad \text{in three dimensions} \qquad (8.31)$$

$$\tau_2 = \left(\frac{R^2}{2D_2}\right)\ln\frac{R}{a} \quad \text{in two dimensions} \qquad (8.32)$$

$$\tau_1 = \left(\frac{R^2}{3D_1}\right) \quad \text{in one dimension} \qquad (8.33)$$

where D_i ($i = 1, 2, 3$) are the diffusion coefficients for the indicated dimensions. These times can be substan-

tially shorter in one and two dimensions than in three. For example, a ligand could be imagined to bind rapidly to a protein in a membrane by first binding rapidly and nonspecifically to the membrane, then diffusing through the plane of the membrane to the protein. Although this is an attractive and popular idea, there are no well-characterized examples of this phenomenon.

One-dimensional systems might not be expected in biological systems, but they are probably approximated by long, linear macromolecules. For example, DNA is a long, extended double-helical structure with a high density of ionized phosphate groups. Specific DNA-binding proteins find short, specific nucleotide sequences in such molecules much more rapidly than would be expected for normal diffusion through solution in three dimensions. Also, the rate of finding specific sequences is observed to be increased by extending the length of the molecule with nonspecific sequences, whereas the opposite might be expected. The proteins appear initially to bind nonspecifically anywhere along the DNA molecule and then to diffuse one-dimensionally along the linear molecule until the correct sequence is found. There may also be "jumping" by the protein molecules between segments of the same DNA molecule that happen to come into proximity in solution. The initial, nonspecific binding of such a protein to DNA is caused by electrostatic interactions between the two and by the release of loosely bound counterions (Sec. 8.3.2). Upon reaching the correct sequence, the binding interactions are much more specific and tighter. Similar considerations may apply to the binding of proteins to long protein aggregates, such as the muscle protein actin.

References

The asymmetric distribution of charges on the surface of horse cytochrome *c*: functional implications. W. H. Koppenol and E. Margoliash. *J. Biol. Chem.* 257:4426–4437 (1982).

Diffusion-controlled macromolecular interactions. O. G. Berg and P. H. von Hippel. *Ann. Rev. Biophys. Biophys. Chem.* 14:131–160 (1985).

Reflections on the kinetics of substrate binding. H. Gutfreund. *Biophys. Chem.* 26:117–121 (1987).

Diffusion effects on rapid bimolecular chemical reactions. J. Keizer. *Chem. Rev.* 87:167–180 (1987).

Real-time spectroscopic analysis of ligand–receptor dynamics. L. A. Sklar. *Ann. Rev. Biophys. Biophys. Chem.* 16:479–506 (1987).

Dynamical simulation of rate constants in protein–ligand interactions. D. A. Case. *Prog. Biophys. Mol. Biol.* 52:39–70 (1988).

Facilitated target location in biological systems. P. H. von Hippel and O. G. Berg. *J. Biol. Chem.* 264:675–678 (1988).

8.2.4 Affinity Chromatography

The interactions of proteins with specific ligands makes possible one of the most powerful methods of protein purification, **affinity chromatography**. The ligand is chemically attached to an insoluble and porous solid support that is suitable for chromatography, in such a way that the ligand is still available for the protein to bind to it. To ensure that the ligand is sufficiently distant from the support that binding of the protein is not physically obstructed, a spacer group is usually inserted between the ligand and the chromatographic support. Obviously, the site on the ligand that is appropriate for its attachment to the solid support is one that is not involved in binding to the protein, but such details of the binding interaction are frequently not known. If the parts of the ligand required for binding to the protein are not known, various linkages between ligand and support must be tried until one is found to work.

In the ideal case, only the desired protein binds to the ligand resin, and all other proteins pass straight through the column. After the column is washed to remove all other proteins, those molecules that are bound tightly to the affinity resin can be eluted by adding soluble ligand to compete with the support or by changing the conditions to decrease the affinity of protein for the bound ligand; in extreme cases, the protein may have to be unfolded before it can be eluted. The use of different concentrations of soluble ligand permits quantitative measurements of the protein's affinities for both the free and the support-bound ligand, which are not necessarily the same. When multiple proteins bind to the same ligand, they can often be eluted separately by different ligands (Fig. 8.7) or by different concentrations of the same ligand (Fig. 8.8).

The principles of affinity chromatography are similar to those of the other types of column chromatography used with proteins, such as ion-exchange, hydrophobic, or reverse-phase chromatography; it differs only in the specificity of the interactions between the protein and the column resin.

References

Affinity chromatography of macromolecules. P. Cuatrecasas. *Adv. Enzymol.* 36:29–89 (1972).

Biospecific affinity chromatography and related methods. J. Porath and T. Kristiansen. In *The Proteins*, 3rd ed., H.

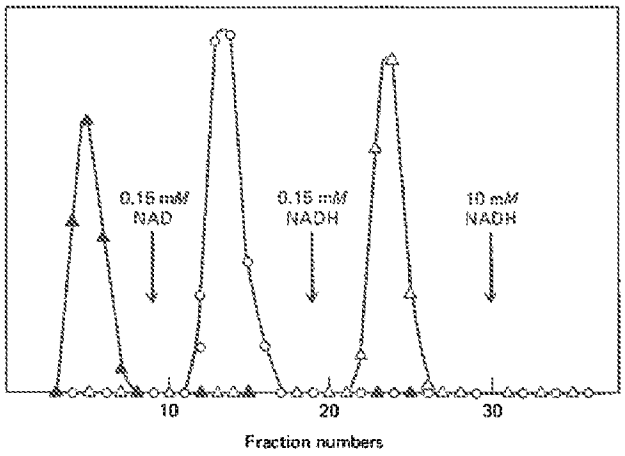


FIGURE 8.7
Separation of three proteins by affinity chromatography. The column matrix was Sepharose with the cofactor NAD linked covalently to it by a diaminohexane spacer. A mixture of bovine serum albumin, glyceraldehyde 3-phosphate dehydrogenase, and lactate dehydrogenase was applied to the column. The two dehydrogenases bind NAD and were bound by the column, whereas serum albumin (▲) does not bind NAD and was not retarded. Lactate dehydrogenase (Δ) binds NADH more tightly than NAD and was eluted by low concentrations of NADH. In contrast, glyceraldehyde-*P* dehydrogenase (○) binds NAD more tightly than NADH and was eluted from the column by low concentrations of free NAD. (Adapted from K. Mosbach et al., *Biochem. J.* 127:625–631, 1972.)

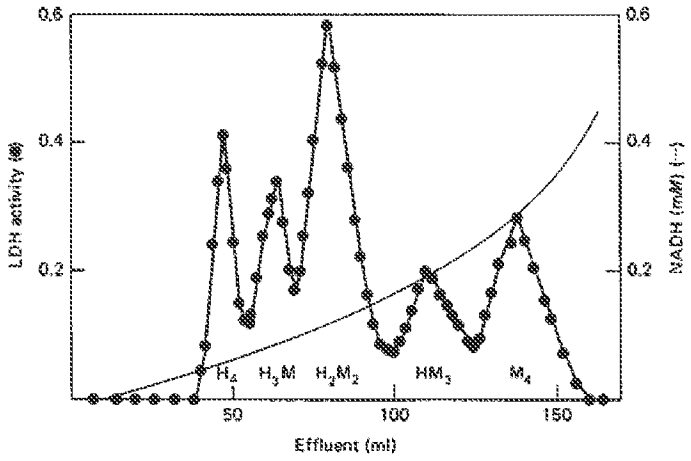


FIGURE 8.8
Separation of the five isozymes of lactate dehydrogenase by affinity chromatography. This enzyme is a tetramer that can be formed from two different but related polypeptide chains, usually designated H and M. The two chains can associate nearly randomly, so five tetramers are possible: the H_4, H_3M, H_2M_2, HM_3, and M_4 isozymes. The individual subunits of the tetramer bind NADH independently, with H chains having a fivefold greater affinity than M chains. Consequently, H_4 is eluted from an affinity column with lower concentrations of free NADH in the buffer and the others in order of their average affinities. The affinity column ligand was AMP, which corresponds to half an NADH molecule, bound to Sepharose. (Adapted from P. Brodelius and K. Mosbach, *FEBS Letters* 35:223, 1973.)

Neurath and R. L. Hill, eds., vol. 1, pp. 95–178. Academic Press, New York, 1975.

The purification of biologically active compounds by affinity chromatography. M. Wilchek and C. S. Hexter. *Methods Biochem. Anal.* 23:345–385 (1976).

The role of protein structure in chromatographic behavior. F. E. Regnier. *Science* 238:319–323 (1987).

Receptor-affinity chromatography. P. Bailon and D. V. Weber. *Nature* 335:839–840 (1988).

8.3 Relationship between Protein Conformation and Binding

Binding sites comprise relatively little of the structure of most proteins and often are just a small patch on the protein surface; small localized alterations of the protein, therefore, can produce large changes in its affinities for ligands, without changes in the overall protein conformation. Consequently, homologous proteins can bind very different ligands. Examples are the bacterial binding proteins present in the periplasm and involved in uptake of nutrients from the medium. All of these proteins that have been studied have very similar three-dimensional structures, consisting of two domains with a single binding site between them. Yet the various proteins bind a variety of small molecules, often with extreme specificity, including monosaccharides, oligosaccharides, amino acids, oligopeptides, sulfate (Fig. 8.4), and phosphate. Other examples are the immunoglobulins (Sec. 8.3.1), in which the same structural framework is used to bind an immense variety of antigens.

Conversely, the same ligand can be bound by unrelated proteins with different three-dimensional structures. This is illustrated most spectacularly by the binding of the heme group (iron-protoporphyrin IX) by a variety of proteins, especially the globins and various cytochromes. The globins (e.g., hemoglobins, myoglobin, erythrocruorin, and leghemoglobin) are all homologous and bind the heme group similarly (see Fig. 6.30). Very different are the structures of a large number of cytochrome *c*–like proteins (see Fig. 6.31) and of the unrelated cytochrome b_5, cytochrome b_{562}, and cytochrome c_3. This last binds four heme groups simultaneously in different ways at four different sites. Thus, a number of proteins appear to have acquired independently the ability to bind heme groups. The ability to bind the same ligand, therefore, cannot be used to imply that two proteins are related.

Nevertheless, the members of a number of different protein families do tend to bind similar ligands. Examples are the globins (Sec. 8.4.3) and the various protein families that bind DNA (Sec. 8.3.2) and Ca^{2+} ions (Sec. 8.3.4.a). The proteins in each of these examples are known to be evolutionarily related, and they obviously have retained their ligand-binding functions. In other cases, an evolutionary relationship is not obvious, and there is the possibility that similar three-dimensional structures have evolved to bind similar ligands in similar ways, perhaps for physical reasons. For example, the numerous 8-fold $\alpha\beta$ barrel proteins (Sec. 6.4.2) are the most likely candidates for examples of convergence to the same three-dimensional structure, yet they all seem to bind negatively charged ligands in similar positions at the same end of the barrel structure. Similarly, nucleotides are bound in different proteins to similar supersecondary structures, known collectively as the nucleotide-binding fold or the Rossmann fold (Sec. 8.3.3), that have no detectable amino acid sequence similarities. Do these similar structures and interactions with ligands reflect common ancestry? Or do they reflect similar physical principles of structure and binding? The verdict is not yet in.

The question of the relationship between protein structure and ligand binding is important because the biological functions of newly discovered proteins are often inferred from their homology to proteins of known structure and function. Rather than just give rules, some of the most informative examples are described.

References

The evolution of dehydrogenases and kinases. W. Eventoff and M. G. Rossmann. *Crit. Rev. Biochem.* 3:111–140 (1975).

Relation between structure and function of α/β proteins. C. I. Brändén. *Quart. Rev. Biophys.* 13:317–338 (1980).

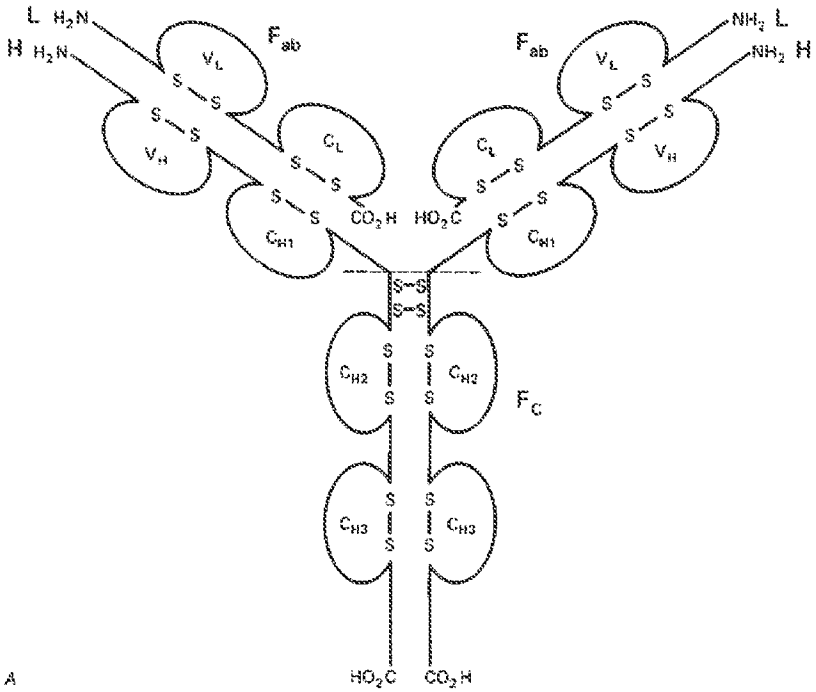
The structural motif of β-lactoglobulin and retinol-binding protein: a basic framework for binding and transport of small hydrophobic molecules? I. Godovac-Zimmerman. *Trends Biochem. Sci.* 13:64–66 (1988).

Crystal structures of two intensely sweet proteins. S. H. Kim et al. *Trends Biochem. Sci.* 13:13–15 (1988).

Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria. F. A. Quiocho. *Phil. Trans. Roy. Soc. Lond. B* 326:341–351 (1990).

8.3.1 Immunoglobulins

Antibody molecules are capable of prodigious diversity; individually they bind a few antigens very specifically,



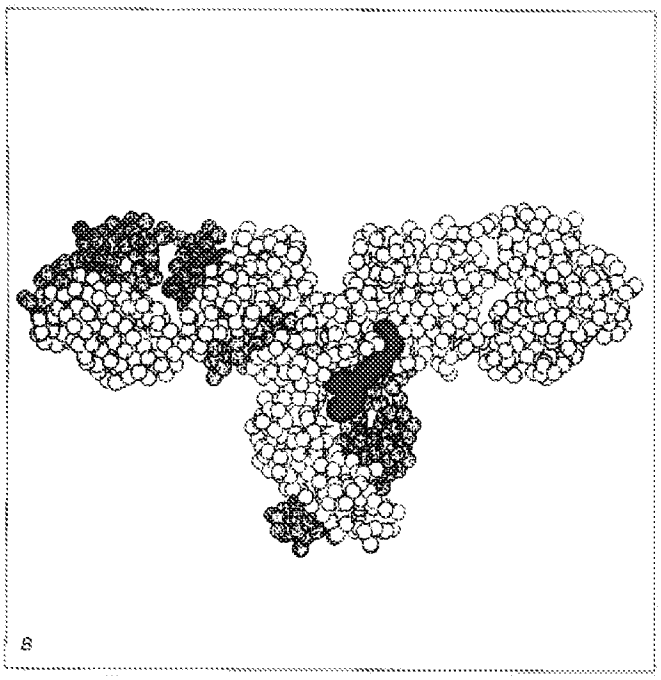


FIGURE 8.9

Schematic representation of a typical immunoglobulin structure *(A)* and a space-filling model determined crystallographically *(B)*. *A* shows the L and H polypeptide chains as solid lines, with the intramolecular disulfides linking Cys residues, about 60 residues apart in the primary structure, that are characteristic of each immunoglobulin domain. The site of cleavage by papain is shown by the dashed line; this cleavage yields two F_{ab} fragments and one F_c. If cleavage occurs on the carboxyl side of the disulfide linking the H chains, as occurs with pepsin, the two F_{ab}-like fragments are linked by the disulfide and are usually designated as F'_{ab}. In *B*, each sphere represents one amino acid residue. One complete heavy chain is white, the other heavily shaded; both light chains are white. The carbohydrate attached to the C_{H2} domain of each heavy chain is black. The antigen-binding sites are at the tips of the F_{ab} arms, at the far left and far right, where the V_H and V_L domains meet. (From E. W. Silverton, et al., *Proc. Natl. Acad. Sci. USA* 74:5140–5144, 1977.)

while collectively they are able to recognize virtually any molecule. Despite this binding diversity, antibody molecules have common structural and functional features. The consequences of ligand binding include elimination of the antigen–antibody complex from the bloodstream, complement-induced lysis of cells, histamine release, and stimulation of secretion of antibodies by lymphocytes, depending on the class of antibody. All immunoglobulin molecules of a given class must have these common functions, which are then combined with different specificities for various antigens. How these different functions are combined is a major question that is only now being answered in molecular terms.

The basic structure of an intact immunoglobulin is a Y-shaped molecule composed of two H and two L chains (Fig. 8.9); in immunoglobulins A and M, these molecules are assembled further into larger complexes. The L chain consists of V_L and C_L domains, and the H chains consist of V_H, C_{H1}, C_{H2}, and C_{H3} domains. All these domains consist of about 100 residues, are homologous in their primary structures, and are independent, stable structural units. All immunoglobulin domains have essentially the same conformation, designated the **immunoglobulin fold**, consisting of two layers of antiparallel β-sheets that are usually linked by a disulfide bond (Fig. 8.10). The arrangement of the various domains in the Y-shaped molecule is shown schematically in Figure 8.9A. The polypeptide chains between the domains are susceptible to proteases. Most susceptible is the hinge region linking the two arms to the base of the Y. After cleavage at this site, the two arms are released individually and are known as F_{ab} fragments; the base is known as the F_c fragment. Each F_{ab} fragment contains the V_L, C_L, V_H, and C_{H1} domains; the F_c fragment has two copies of each of the C_{H2} and C_{H3} domains. F_V fragments are produced in other ways and consist of only the V_H and V_L domains. The connecting segments between the domains have varying degrees of flexibility, and the individual domains undergo considerable motion relative to each other. The Y-shaped molecule shown in Figure 8.9 is just one of many shapes that an immunoglobulin can adopt in solution.

The immunoglobulin domains interact with each other in a variety of ways in an intact H_2L_2 molecule (Fig. 8.9B). The two C_{H2} domains interact with each other in the F_c portion, as do the two C_{H3} domains. Each arm of the molecule is composed of one C_{H1} domain interacting with one C_L domain, plus the interacting V_H and V_L domains. All pairs of C domains associate in a similar manner in which the members of one of their pairs of β-sheets associate isologously (see Fig. 6.25); in contrast, the V domains associate by means of the other β-sheet. Less extensive, and presumably less stable, interactions also take place between domains adjacent in the primary structure. Nevertheless, the segments of polypeptide chain linking the domains have varying degrees of flexibility; especially flexible is the hinge region linking the F_{ab} and F_c portions, which makes the two F_{ab} arms particularly mobile.

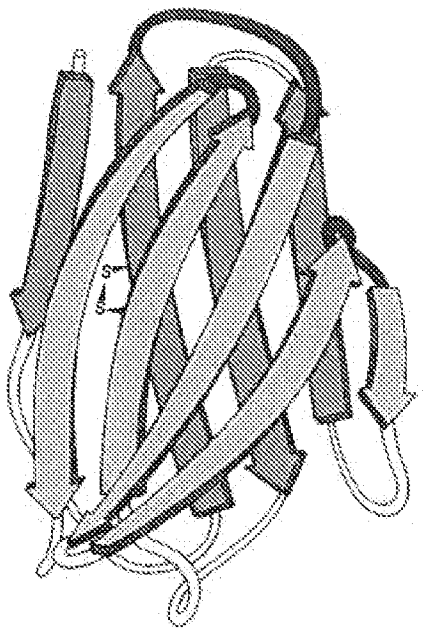


FIGURE 8.10
The immunoglobulin fold. Two layers of antiparallel β-sheet are folded on top of each other to form a sandwichlike structure. Between the two layers are hydrophobic side chains and the indicated disulfide bond linking two Cys residues about 60 residues apart; one Cys residue is in the middle of the second strand from the left of the top sheet; the other is in the middle of the second strand of the bottom sheet. In V_L domains, as illustrated here, the top β-sheet has five strands, the bottom four. The two strands at the right edge of the five-strand sheet are missing in C domains. The loops containing the hypervariable regions in V domains are shaded dark. (From J. Richardson. *Adv. Protein Chem.* 34:167–339, 1981.)

References

Three-dimensional structure of antibodies. P. M. Alzari et al. *Ann. Rev. Immunol.* 6:555–580 (1988).

Elbow motion in the immunoglobulins involves a molecular ball-and-socket joint. A. M. Lesk and C. Chothia. *Nature* 335:188–190 (1988).

Antibody: the flexible adaptor molecule. D. R. Burton. *Trends Biochem. Sci.* 15:64–69 (1990).

a. Recognition of Antigens

The antigen-binding sites are at the tips of the two F_{ab} arms, situated between the V_L and V_H domains. Both F_{ab} and F_V fragments retain the antigen-binding sites. Each site is composed of the residues of the three irregular loops between β-strands of both the L and H chains (Fig. 8.11). Different binding sites are generated with different amino acid side chains in these positions, which are known as the **complementarity determining regions**, or CDRs. Their conformations depend largely on the conserved structure of the remainder of the immunoglobulin fold, which serves as a scaffold (Fig. 8.10). Consequently, the conformations of the CDR regions of many antibodies can be predicted with some degree of success from just their amino acid sequences. Also, the CDR regions from one immunoglobulin can be grafted onto the scaffold of another, and it is now becoming possible to design antibodies to order.

An enormous number of antibody molecules with different antigen specificities are made by complex organisms, and this diversity of immunoglobulin primary structures is generated by a special mechanism during their biosynthesis. The variable domains are encoded by separate gene segments, designated as variable (V_H), diversity (D), and joining (J) segments for the heavy chain and V_L and J_L segments for the light chain. For the heavy chain in the mouse, there are 100–1000 V_H elements, approximately 12 D elements, and 4 J elements. Similarly, the light chain is encoded by more than 100 V and 5 J_L elements. Different antibody molecules are generated by joining these elements in different combinations. Further genetic variation is introduced at the sites where these gene segments are joined together by the genetic fusion mechanism. In this way, 10^8–10^{10} different antibody molecules can be generated from a limited number of gene segments. Most of this variation is produced in the residues comprising the three CDR regions of each polypeptide chain. Accordingly, the CDR regions are hypervariable in different antibody molecules, and they are also known as the **hypervariable regions**.

Each antibody-producing cell normally produces a single antibody molecule. When such a cell encounters an antigen that its antibody recognizes sufficiently (at this stage known as an *immunogen*), the cell is induced to synthesize the antibody in large quantities, to undergo cell division, and to proliferate. In the process, the genetic segment coding for the variable region of the antibody molecule undergoes mutation at a rate much greater than normal. Progeny cells producing antibodies with greater affinity for the immunogen are then selected. Consequently, the immunological response to an immunogen changes with time. Initially, many low-affinity antibodies are produced, with $K_d \approx 10^{-5}$–10^{-7} M, but with time, antibodies of increasing affinity are produced. In an organism, many different antibody molecules are produced by many different cells, so normal antisera are very heterogeneous. Individual antibody-producing cells can be selected and cloned, however, and homogeneous **monoclonal antibodies** produced by them. Furthermore, the methods of protein engineering (Sec. 2.2) make it possible to manipulate the genes for antibodies, to express them in microorganisms, and to select for antibodies of the desired specificity.

There is one important restriction in the number of different antibody molecules that the normal immune system produces naturally against various antigens. Antibodies are not usually produced against molecules that are normally present in the host organism, for obvious reasons.

The interactions of antibody-combining sites with antigens seem to be no different from the interactions of other proteins with ligands. They occur with comparable rates and binding constants. Crystal structures of several immunoglobulin–antigen complexes demonstrate the usual, but still remarkable, complementarity of shapes of the CDR regions and of the antigen, with the usual van der Waals interactions and hydrogen bonds. No large conformational changes take place in the antibody molecules upon binding antigens. The six CDR loops play varying roles in the various complexes, and not all necessarily interact with each antigen. Generally, the heavy-chain CDR regions seem to play the major role, and single V_H domains often have significant affinities for ligands.

A major question to be answered is how the binding of antigens to the binding sites of intact immunoglobulins at the tips of the F_{ab} arms triggers the effector functions, such as complement activation, which are a property of the F_c portion of the molecule. Different classes of immunoglobulins have different C domains and different effector functions. There is little evidence for specific conformational changes in immunoglobulins upon antigen binding, although the immunoglobulins undergo a wide variety of motions of the various domains relative to each other. Effector functions appear to be triggered primarily by the formation of large antigen-immunoglobulin aggregates. In particular, the first protein of the classical complement activation pathway, C1q, is a complex structure resembling a bunch of six tulips, with six globular heads joined by six collagen-like stems that are held together in the lower half. Each head binds to the C_{H2} domain of an immunoglobulin; the simultaneous binding of many heads to aggregated immunoglobulins and antigens appears to be the trigger that sets off complement activation.

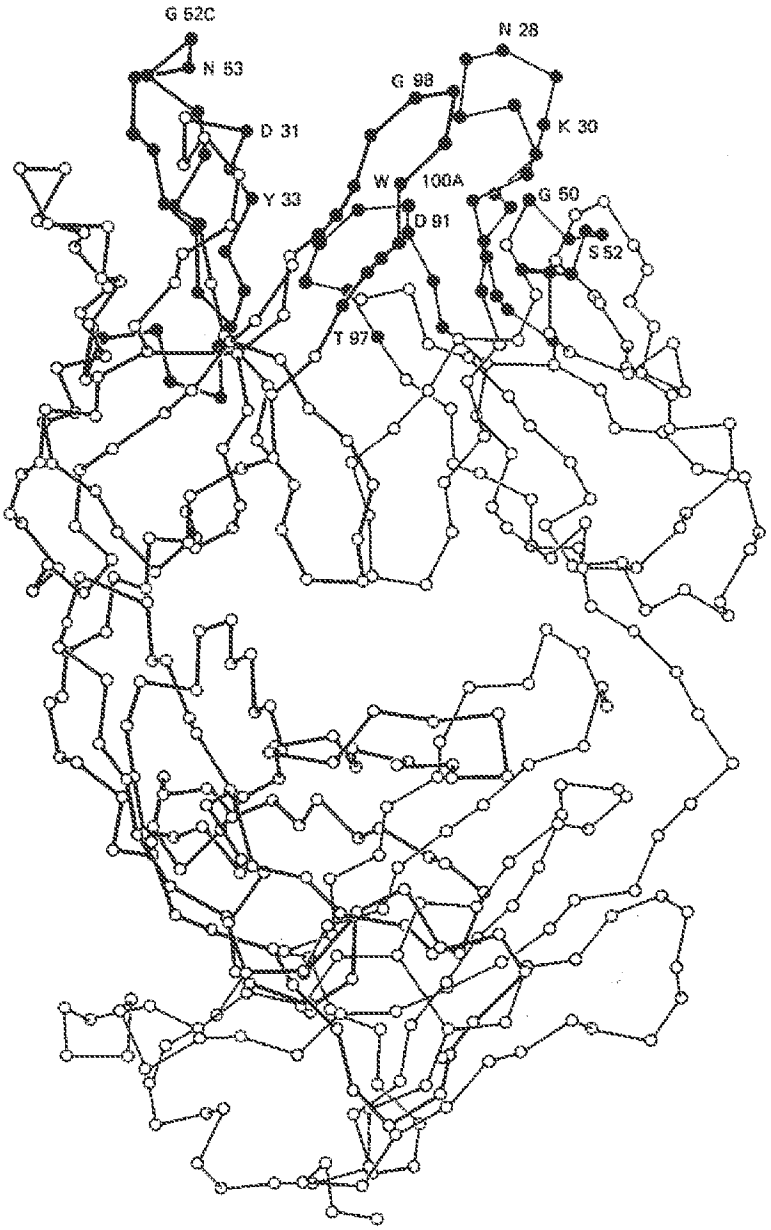


FIGURE 8.11
The α-carbon backbone of the F_{ab} fragment of immunoglobulin M_{60S}. The hypervariable residues comprising the antigen-binding site are darkened. Two residues of each hypervariable segment are identified using the one-letter code for the amino acids. The heavy chain is shown with thicker lines than the light chain. The top half of the structure comprises the V_H and V_L domains, the bottom half the C_L and C_{H1} domains. (Kindly provided by D. R. Davies and H. Metzger.)

References

Thermodynamics of hapten–antibody interactions. T. K. S. Mukkur. *Crit. Rev. Biochem.* 16:133–167 (1987).

Diversity and the genesis of high affinity antibodies. C. Milstein. *Biochem. Soc. Trans.* 15:779–787 (1987).

Reshaping human antibodies for therapy. L. Riechmann et al. *Nature* 332:323–327 (1988).

Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. E. S. Ward et al. *Nature* 341:544–546 (1989).

Conformations of immunoglobulin hypervariable domains. C. Chothia et al. *Nature* 342:877–883 (1989).

The role of somatic hypermutation in the generation of antibody diversity. D. L. French et al. *Science* 244:1152–1157 (1989).

Antibodies: a paradigm for the biology of molecular recognition. C. Milstein. *Proc. Roy. Soc. Lond. B.* 239:1–16 (1990).

Antibody–antigen complexes. D. R. Davies et al. *Ann. Rev. Biochem.* 59:439–473 (1990).

Three-dimensional structure determination of an anti-2-phenyloxazolone antibody: the role of somatic mutation and heavy/light chain pairing in the maturation of an immune response. P. M. Alzari et al. *EMBO J.* 9:3807–3814 (1990).

Man-made antibodies. G. Winter and C. Milstein. *Nature* 349:293–299 (1991).

b. Proteins as Antigens

If no ligands are known for a given protein, they can be made to order by preparing antibodies against it. This can be extremely important for identifying and quantifying a protein. For example, a new protein known only from its gene sequence can usually be identified by synthesizing a peptide corresponding to a short segment of its primary structure, preparing antibodies against this peptide, and using these antibodies to identify the protein. This procedure is most successful if this peptide corresponds to a part of the primary structure that is accessible to antibodies, even in the folded native conformation.

Specific proteins are often identified immunochemically by a procedure known as **Western blotting**. Proteins in samples separated by electrophoresis in polyacrylamide gels (Sec. 1.5.3) are transferred electrophoretically, in a way that retains the electrophoretic pattern, to a membrane to which the proteins stick tightly. The membrane is then treated with antibodies specific for the protein to be detected, and the positions of the bound antibodies are visualized in various ways. Using Western blotting, a single protein can be detected specifically among a mass of other proteins. The technique of Western blotting need not employ antibodies but can use other ligands that bind specifically to certain proteins. At least some proteins seem, perhaps surprisingly, to refold after or during transfer to the membrane because they exhibit ligand-binding properties that are dependent on the folded conformation of the protein.

There are numerous other immunochemical methods for quantifying specific proteins, some of which are of clinical importance. All of these immunochemical procedures are dependent on the conformational specificities of antibodies directed against proteins and peptides, and their cross-reactivities, so there has been substantial interest in proteins as immunogens. After much controversy, the determination of several crystal structures of complexes of proteins bound to antibodies directed against them (Fig. 8.12) has substantially clarified the situation. These crystal structures demonstrate that neither the antibody nor the protein antigen change their conformations substantially upon interacting. The interface between them is similar to those between other interacting protein molecules. Antibodies generally recognize an area on the surface of the protein determined largely by the size of the binding site on the antibody; this binding invariably involves atoms from residues of the protein that are distant in the sequence but are brought into close proximity by the protein's conformation. All antibodies against proteins are specific in some way for their conformations, which is the basis of their use in measuring conformational equilibria (Sec. 7.5.3.d).

Native proteins used as immunogens generally elicit at least some antibodies specific for the folded conformation, depending on the stability of the protein to the immunization procedure. Antibodies recognizing the folded conformation may be directed against any portion of the protein surface unless the host organism contains a protein with the same surface. If the protein is not very stable, unfolded molecules will also be present or will be produced by the immunogenic manipulations, and antibodies recognizing the unfolded state will also be produced. Proteins unfolded irreversibly and peptide fragments of a protein can also be used to produce antibodies recognizing the unfolded state. The specificity of such antibodies remains to be determined, but it is most likely that each antibody is specific for one of the many conformations that disordered polypeptides can adopt. Accordingly, unfolded proteins and peptides have substantially lower affinities for their antibodies than do folded proteins.

Antibodies against folded proteins cross-react with unfolded proteins or peptide fragments, and vice versa, to varying extents, depending on the conformational flexibilities of the two conformational states of the protein (Sec. 7.5.3.d). Antibodies against a peptide frag-

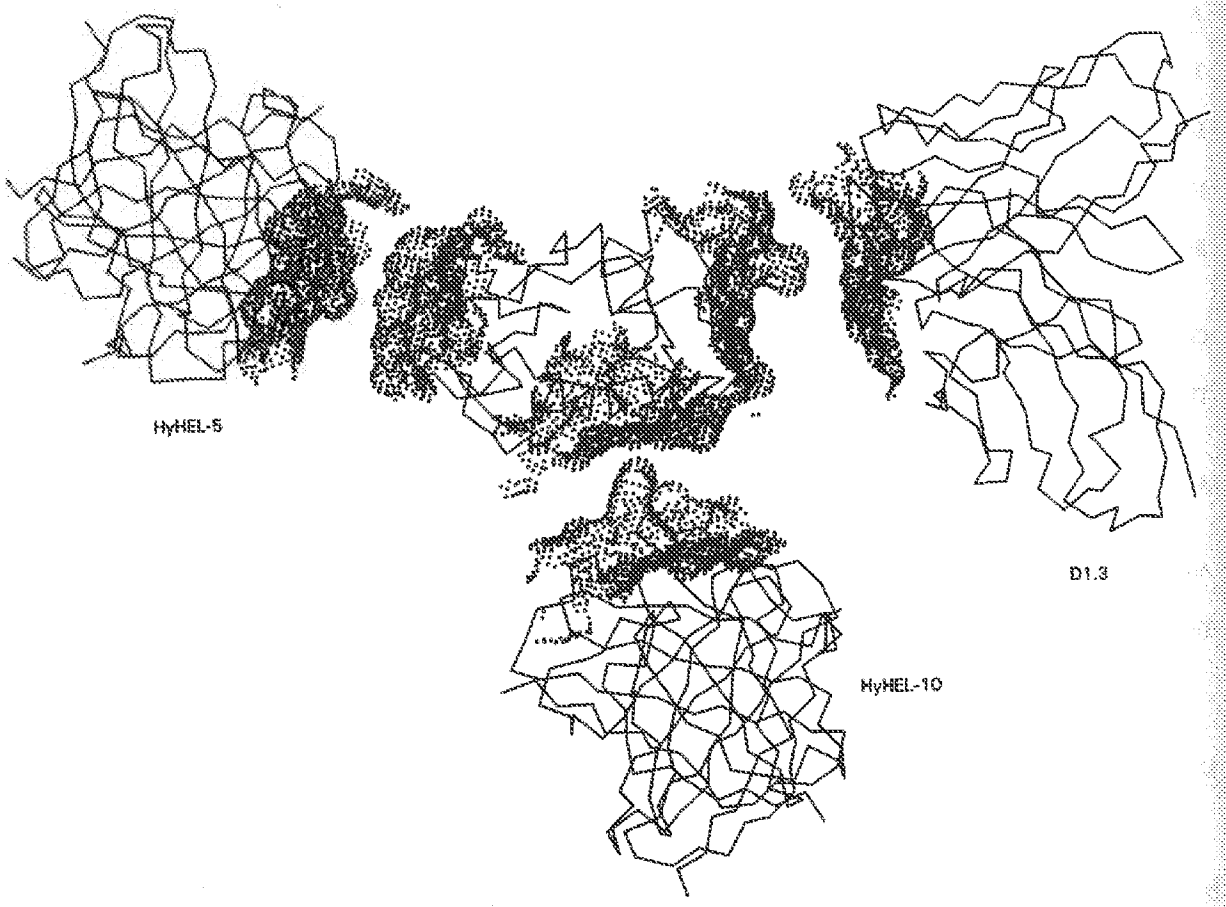


FIGURE 8.12
A composite of hen lysozyme binding to the F_V regions of three different antilysozyme antibodies, D1.3, HyHEL-5, and HyHEL-10. For simplicity of illustration, the three F_V fragments have been pulled away somewhat from the lysozyme molecule in the center. C^α representations are used for the F_V fragments and for lysozyme, and a dot representation is used for their interacting surfaces. Note that the three epitopes on lysozyme recognized by these three antibodies do not overlap, except for a small overlap between those of HyHEL-10 and D1.3. (From D. R. Davies et al., *Ann. Rev. Biochem.* 59:439–473, 1990.)

ment recognize the native protein to the extent that it unfolds spontaneously or if it is induced to do so by the manipulations used in the immunochemical measurements. The possibility that the peptide fragment tends to generate antibodies against a particular conformation must also be considered.

Antibody molecules can also be used, in a different species, as immunogens. Of special interest are the antibodies that are directed against their antigen-binding regions, known as **anti-idiotopes**. If such anti-idiotopes are complementary to the immunogenic antibody binding site, they should be equivalent to the antigen to which the immunogenic antibody was raised (Fig. 8.13). Anti-idiotopes should be useful for binding studies in place of that antigen, and this expectation has been fulfilled in some instances. Although the antigen and anti-idiotope surfaces are equivalent to at least some extent, sufficient to permit their binding to the same antibody molecule, they need not be structurally equivalent.

References

The antigenic structure of proteins: a reappraisal. D. C. Benjamin et al. *Ann. Rev. Immunol.* 2:67–101 (1984).

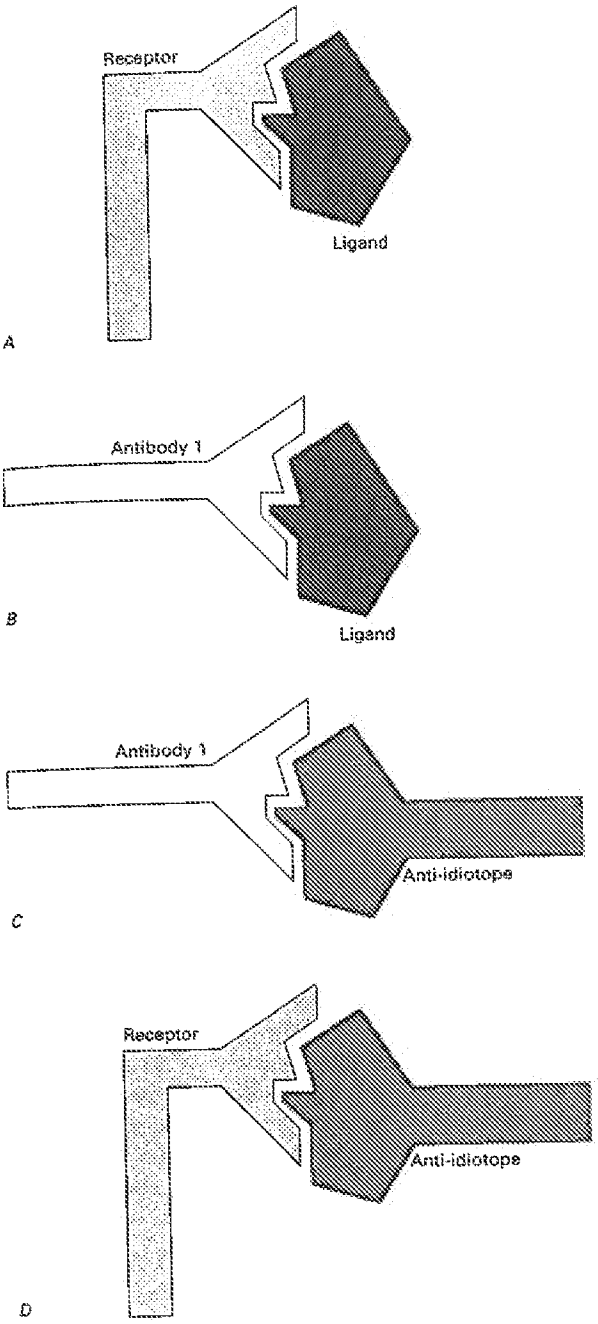


FIGURE 8.13
Schematic illustration of anti-idiotope antibodies and their uses. *A:* A sought-after receptor for a specific ligand bound to the ligand. *B:* Antibodies directed against the functional region of the ligand should have a combining site that mimics the receptor. *C:* An anti-idiotopic antibody that is raised against the combining site of the first antibody should have a combining site that mimics the ligand. *D:* Such anti-idiotopes should be able to identify the sought-after receptor.

Continuous and discontinuous protein antigenic determinants. D. J. Barlow et al. *Nature* 322:747–748 (1986).

The structural basis of antigen–antibody recognition. R. A. Mariuzza et al. *Ann. Rev. Biophys. Biophys. Chem.* 16:139–159 (1987).

Antibody–antigen complexes. D. R. Davies et al. *J. Biol. Chem.* 263:10541–10544 (1988).

On the attribution of binding energy in antigen–antibody complexes McPC 603, D1.3, and HyHEL-5. J. Novotny et al. *Biochemistry* 28:4735–4749 (1989).

Antibody geometry and form: three-dimensional relationships between anti-idiotypic antibodies and external antigens. W. V. Williams et al. *Trends Biotechnol.* 8:256–263 (1990).

Three-dimensional structure of an idiotope–anti-idiotope complex. G. A. Bentley et al. *Nature* 348:254–257 (1990).

8.3.2 *DNA-Binding Proteins*

Proteins that bind specifically to DNA are of great biological importance because they are usually involved in replication or expression of genetic information. Of greatest interest have been those proteins that bind to very specific sites on the DNA, defined by specific sequences of the four nucleotides A, T, C, and G at a few adjacent positions. For a protein to distinguish among different nucleotide sequences in double-stranded DNA is not straightforward because the nucleotides of the two antiparallel strands are base-paired (A-T, T-A, C-G, and G-C) in the interior of the double helix (Fig. 8.14). The exterior surface of the double helix is almost independent of its nucleotide sequence, being composed primarily of the constant phosphate–sugar backbone. Only the edges of the nucleotides are accessible to the solvent and to the protein, primarily in the major groove of the DNA double helix. The nucleotides are distinguished primarily by the polar groups that are accessible.

If a protein is to discriminate among DNA base-pairs by interacting with their edges in the major groove, it needs to have interacting groups that protrude substantially from its surface, to be able to contact the nucleotides at the base of the groove. The best characterized structural motif that accomplishes this is the **helix-turn-helix**, which protrudes from the protein surface. It is observed in a number of proteins that have no other structural similarities (Fig. 8.14). This structural motif seems to have sufficient intrinsic stability to be able to exist as a protuberance, with few interactions with the rest of the protein structure, in order to penetrate the DNA major groove. A number of hydrophobic